From the Richness of the Signal to the Poverty of the Stimulus: Mechanisms of Early Language Acquisition

Judit Gervain

2007
Trieste

Contents

List of Figures

List of Tables

List of Most Frequently Used Abbreviations

ANOVA:	Analysis of Variance
AOI:	Area of Interest
AP:	Adjectival Phrase
BW TP:	Backward Transitional Probability
C:	Consonant
Comp:	Complement
deoxyHb:	Deoxygenated Hemoglobin
DP:	Determiner Phrase
EEG:	Electroencephalography
ERP:	Event Related Potential
(f)MRI:	(functional) Magnetic Resonance Imaging
F(X):	Frequency of unit X (in a linguistic sample)
FW TP:	Forward Transitional Probability
FW:	Frequent Word
IW:	Infrequent Word
LH:	Left Hemisphere
MB:	Morpheme Boundary
MDL:	Minimum Description Length (algorithm)
MI:	Mutual Information
N:	Noun
NIRS:	Near-Infrared Spectroscopy
NP:	Noun Phrase
O:	Object
OT:	Optical Topography
oxyHb:	Oxygenated Hemoglobin
P(X):	Probability of unit X (in a linguistic sample)
P&P (theory):	Principles and Parameters (theory)
PP:	Prepositional Phrase
RH:	Right Hemisphere
ROI:	Region of Interest

S (in transformational rules):	Sentence
S (in word order typology):	Subject
Spec:	Specifier
TMS:	Transcranial Magnetic Stimulation
TP:	Transitional Probability
UG:	Universal Grammar
V (lexical category):	Verb
V (phoneme category):	Vowel
VP:	Verb Phrase
WB:	Word Boundary
WI:	Word Internal (Syllabic Transition)
XP:	X (i.e. some undefined) Phrase

Abbreviations used in morphemic glosses of linguistic examples are explained in separate footnotes at the first occurrences of the abbreviations.

Acknowledgments

First and foremost, I would like to thank Jacques Mehler for initiating me into developmental psychology. His enthusiasm about baby work, optical topography and language acquisition has been a great source of inspiration. I am grateful that he has been open to weird linguistic notions, such as the Head-Complement parameter or agglutination. Many thanks to Marina Nespor for pushing the limits of this openness even further. Thanks to Marina and Jacques for all the good time: the delicious dinners, the good laughs, the beautiful view from their place, the trips to Hungary . . .

I am grateful to all the members of the Language, Cognition and Development Lab at SISSA for all that I have learned from them, for the lively discussions during the lab meetings and for all the fun. Thanks go to Luca Bonatti for his firm ideological stance, for teaching me so much about PsyScope and Macs and for having those wonderful kids. I am grateful to Marcela Peña for her help with the optical topography work. Many thanks to Mohinish Shukla, Ágnes Kovács, Ernő Téglás and Erika Marchetto for all I have learned from them about experiments, programming and babies. I am grateful to Ansgar Endress, Juan Manuel Toro, Damir Kovačić, Jean-Rémy Hochmann, Alan Langus, Silvia Benavides, Luca Coletti, Mahan Azadpour and Katya Vinnik for the discussions and all the help. I would like to thank Alessio Isaja and Luca Filippin for their technical support, as well as Jeanne Moussu von Heimendahl and Marijana Sjekloća for recruiting adult and infant subjects.

While conducting the research presented in the thesis, I was hosted by a number of labs, which generously allowed me to run experiments at their facilities or even ran the experiments for me. I am grateful to Núria Sebastian Gallés and the members of the Cognitive Neuroscience Research Group at the University of Barcelona, who ran several adult experiments for me, and even took the trouble to hunt for Basque speakers in remote villages of the Basque country. Thanks to Csaba Pléh and the Department of Cognitive Sciences at the University of Technology and Economics of Budapest, where I tested all my Hungarian subjects. The Institute of Linguistics of the Hungarian Academy of Sciences gave me the opportunity to present my work at different stages of its development. Thanks to István Kenesei, Zoltán Bánréti and the different audiences for the invitations and the useful feedback I received at the Institute. Emmanuel Dupoux, Franck Ramus and Gayaneh Szenkovits from the Laboratoire de Sciences Cognitives et Psycholinguistique at CNRS-EHESS-ENS welcomed me several times in Paris. I tested my Japanese infants at the Laboratory of Language Development, where I received invaluable help from all the members, in particular Reiko Mazuka and Ryota Horie. All the newborns were tested at the Azienda Ospedaliero-Universitaria "Santa Maria della Misericordia" di Udine. I am indebted to Francesco Macagno and Silvia Cogoi for their help in recruiting and testing the babies.

The work presented in this thesis has benefitted from discussions with a number of people during the different stages of its preparation. I am grateful to all of them. In particular, I would like to thank Richard Aslin, Gergő Csibra, Ghislaine Dehaene-Lambertz, Emmanuel Dupoux, Angela Friederici, György Gergely, Lila and Henry Gleitman, Ryota Horie, István Kenesei, Pino Longobardi, Gary Marcus, Reiko Mazuka, Hellmuth Obrig, Franck Ramus, Csaba Pléh, Luigi Rizzi, Alessandro Treves, Tim Shallice, Gayaneh Szenkovits, Virginia Valian, Janet Werker and Gábor Zemplén.

My wonderful family bore with me during the five years of my absence. In particular, I would like to thank my parents, Frecska Irén and Gervain Mihály, my sister, Gervain Eszter, my grandparents, Frecska András and Frecska Andrásné, my aunt, Frecska Éva and my cousins, Jenei Zsuzsanna and Zubornyák Attila, who visited me several time in Trieste.

I am grateful to all my friends, who made my life in Trieste happy and enjoyable. Thanks to Gábor Zemplén for his company. Natalia Pietrosemoli, Michèle de la Chevrotière, Reinel Sospedra and Danijela Roža were wonderful flatmates, who created a cosy, family atmosphere.

Very special thanks to Ramón for his love, his patience and a lot more.

Chapter 1

Introduction: The Logical Problem of Language Acquisition

In 1637, René Descartes argued that language is unique to humans because only the human mind allows the productive combination of words into meaningful expressions. This, he claimed, was not a matter of articulation or speech, which some humans lack, and some animals possess, but a fundamental fact about the nature of the human mind.

> “Car c’est une chose bien remarquable, qu’il n’y a point d’hommes si hébétés et si stupides, sans en excepter même les insensés, qu’ils ne soient capables d’arranger ensemble diverses paroles, et d’en composer un discours par lequel ils fassent entendre leurs pensées ; et qu’au contraire il n’y a point d’animal tant parfait et tant heureusement né qu’il puisse être, qui fasse le semblable. Ce qui n’arrive pas de ce qu’ils ont faute d’organes, car on voit que les pies et les perroquets peuvent proférer des paroles ainsi que nous, et toutefois ne peuvent parler ainsi que nous, c’est-à-dire, en témoignant qu’ils pensent ce qu’ils disent ; au lieu que les hommes qui, étant nés sourds et muets, sont privés des organes qui servent aux autres pour parler, autant ou plus que les bêtes,

> ont coutume d'inventer d'eux-mêmes quelques signes, par lesquels ils se font entendre à ceux qui, étant ordinairement avec eux, ont loisir d'apprendre leur langue. Et ceci ne témoigne pas seulement que les bêtes ont moins de raison que les hommes, mais qu'elles n'en ont point du tout. Car on voit qu'il n'en faut que fort peu pour savoir parler ; et d'autant qu'on remarque de l'inégalité entre les animaux d'une même espèce, aussi bien qu'entre les hommes, et que les uns sont plus aisés à dresser que les autres, il n'est pas croyable qu'un singe ou un perroquet, qui serait des plus parfaits de son espèce, n'égalât en cela un enfant des plus stupides, ou du moins un enfant qui aurait le cerveau troublé, si leur âme n'était d'une nature du tout différente de la nôtre." (Descartes, 1637/1991, Part 5)

Several centuries later, researchers have still not answered the question "what (if anything) is qualitatively new" (Hauser, Chomsky, & Fitch, 2002) in human language and mind.

> "Most current commentators agree that, although bees dance, birds sing, and chimpanzees grunt, these systems of communication differ qualitatively from human language. In particular, animal communication systems lack the rich expressive and open-ended power of human language (based on humans' capacity for recursion). [. . .] There is, however, an emerging consensus that, although humans and animals share a diversity of important computational and perceptual resources, there has been substantial evolutionary remodeling since we diverged from a common ancestor some 6 million years ago. The empirical challenge is to determine what was inherited unchanged from this common ancestor, what has been subjected to minor modifications, and what (if anything) is qualitatively new." (Hauser, Chomsky, & Fitch, 2002)

The most general formulation of the question then is *what is/are the computational component(s) of the human mind that cause(s) the qualitative differ-*

ence between human language, a discretely infinite combinatorial system, and animal communication, a closed system with finite combinatorics? Finding this/these components(s) requires exploring the mechanisms contributing to language, understanding how they interface with each other and with other components of the mind, and identifying those that are specifically dedicated to the discrete infinity of human language.

Several approaches have been taken to answer these queries. One line of research has looked at the philogenesis of our species, and of language in particular, comparing it to the abilities of other species in order to identify the components that are uniquely human, and might thus underlie the observed difference between humans and other animals. This evolutionary approach has its inherent difficulties (lack of fossil records, the difficulty of interpretation of analogies and homologies between species etc.). However, it is certainly an insightful and increasingly productive avenue of research (Bickerton, 1990; Pinker & Bloom, 1990; Hauser, Chomsky, & Fitch, 2002; Pinker & Jackendoff, 2005; Fitch, Hauser, & Chomsky, 2005).

Another approach has investigated the ontogenesis of our species, and of language in particular, exploring the abilities of humans at different stages of their development and individuating computational components on the basis of their different development trajectories. The present thesis endorses this approach by investigating the initial state of development, more specifically the linguistic abilities of human infants. These early abilities are of special relevance, since they provide insight into the toolbox with which the mind comes equipped to start learning about the environment.

In his seminal work, Lenneberg (1967) provided a systematic exploration of the hypothesis that language is rooted in our biological endowment. Mainly on the basis of data from brain damaged patients, he formulated the critical period hypothesis, according to which there is a biologically determined window of opportunity allowing language to be acquired natively. Once this period is over, language learning abilities degrade, and the resulting acquisition will not be native-like. Although some details of this proposal did not receive empirical support, the idea that language acquisition has its sensitive, even if not critical, period is now well established. A large body of evidence has accumulated, clearly establishing that infants and children are better first and second language learners than adults (e.g. Newport, 1990; Jenkins, 2004). It is not entirely clear what the neurodevelopmental basis of this advantage is. Some researchers attribute it to the maturational decline of the language acquisition faculty (Chomsky, 1965, 2000). Others (e.g. Newport, 1990) argue that young children's advantage in language be related to their disadvantage in other cognitive domains. Their limited memory span and combinatorial skills might not allow them to represent and store large units or long sequences of linguistic input. They are thus constrained to encode only smaller chucks. Given the compositional nature of language, this limitation might lead to an advantage, provided that at least some of the stored chunks correspond to real linguistic constituents, e.g. morphemes etc. Under this view, the task of analytically decomposing the input stream, essential for language acquisition, is facilitated for infants by their limited storage and representational abilities. Adults, in contrast, store larger sequences, which places the full burden of the decomposition task on them. Whatever the ultimate explanation of infants' linguistic advantage might turn out to be, it is undeniable that infants and young children constitute a population of special interest, since they provide a window into how a complex human cognitive ability, language, reaches its mature state. The study of how infants acquire language sheds light not only on the learning mechanisms themselves, but also on the computational abilities that adults possess.

Some such components or learning mechanisms have already been proposed. Thus certain aspects of grammar, especially syntax, are believed to be governed by abstract, symbolic rules containing variables (Chomsky, 1957; Pinker, 1984; Guasti, 2002). For example, a sentence can be described as consisting of a subject noun phrase and a predicate verb phrase (formally, S → NP VP), where noun phrase (NP)[1] and verb phrase (VP) are variables that can take an infinite number of different values (for instance, NP: *the girl*, *the cute girl*, *the cute girl in the pretty skirt*, *the cute girl in the pretty blue skirt*, *the cute girl in the pretty blue skirt that her mother bought at the market that I visited with a friend who* ...). Another component that has been identified

[1] For abbreviations, see p. 1.

(Saffran, Aslin, & Newport, 1996; Tomasello, 2000) is a mechanism tracking statistical information such as frequency of occurrence or (conditional) probability between individual items (words, syllables, phonemes etc.).

Although the existence of rule extraction and statistical learning is well established, it is not clear what the division of labor between them might be during language acquisition. In particular, it is heatedly debated whether the discrete infinity of grammar is the product of the symbolic rule system or whether it can be accounted for by statistical learning alone. The two possibilities have radically different implications. The rule-based component is a mechanism specifically dedicated to language and, to the best of our knowledge, it is unique to humans.[2] Statistical learning mechanisms, on the other hand, operate domain-generally over a vast range of possible inputs from auditory to visual (Fiser & Aslin, 2002), and at least some of their simple forms can be found in nonhuman animals (Hauser, Newport, & Aslin, 2001; Toro & Trobalon, 2005). Under this view, language is more of a quantitative than a qualitative difference between humans and other animals.

A third computational component has recently been proposed to play a role in language acquisition (Endress et al., 2005; Endress, Dehaene-Lambertz, & Mehler, in press). Perceptual 'primitives' or Gestalt-like configurations derive from the architecture of the perceptual system, and might account for why certain patterns are more readily learnable than others. Such mechanisms, while well known in vision research (Hubel & Wiesel, 1959; Gilbert & Wiesel, 1990), have largely been neglected in language acquisition.

When investigating the mechanisms underlying acquisition, it is also revealing to evaluate the input that learners receive. This can provide direct evidence about the kinds of information and representations that learning mechanisms operate on, yielding further tools to distinguish computationally different components of the language acquisition faculty.

[2] While cotton-top tamarins are able to discriminate the simple repetition-based grammars used in Marcus, Vijayan, Bandi Rao, and Vishton's (1999) study (Hauser, Weiss, & Marcus, 2002), it has been shown that perceptual mechanisms focusing on some salient aspects of the stimuli are enough to learn these grammars—rule learning is not necessarily required (Endress, Scholl, & Mehler, 2005).

The present thesis considers language acquisition as a complex learning procedure, in which the three key components, viz. the initial state, the input and learning, place mutual constraints on one another. For instance, the informational content of the input defines computational boundary conditions for the representations that learning mechanisms might use. Therefore, in addition to investigating the initial state of the language faculty, the input that learners receive and the learning that takes place, the thesis also attempts to explore how these components interact during the earliest stages of acquisition.

1.1 The poverty of stimulus argument and the learnability of language

1.1.1 The induction problem

Acquiring language poses a serious learning problem. While infants receive a large amount of exposure to the ambient language, this input contains little explicit information about how it is structured. In the absence of such information, an infinite number of possible rule sets can be found that correctly describe the input data set. Consequently, no successful learning can take place. Note that even if a learner can find a way to settle upon one of the infinitely many rule sets, there is no guarantee that this will converge with the grammars chosen by the other members of the linguistic community. This classical induction problem, known in linguistics as the 'poverty of stimulus' argument, was first emphasized and applied to language acquisition by Noam Chomsky (1957, 1959). He drew a parallel between the explicit task of a linguist who is attempting to describe the grammar of an unknown language, and the implicit endeavor of the language learning infant, who is discovering the grammar of his or her native language.[3]

[3] It is interesting to note from a historical perspective that reflections on the induction problem in language are rooted in the works of North American structuralist linguists, such as Leonard Bloomfield (1933, 1939) and Zellig Harris (1951, 1955), who, unlike their European colleagues, where faced with the problem of having to analyze previously

Importantly, the poverty of stimulus problem is further aggravated by the fact that infants only hear possible sentences. They receive no *negative information*, that is, information about what is *not* a possible sentence of the target language. Therefore, they have no way to actively test and eliminate at least some of the competing grammar candidates. Indeed, it has formally been shown that languages as complex as human grammars cannot be learned from the input in the absence of negative evidence (M. E. Gold, 1967); although under special conditions, learning has been claimed to be possible on the basis of positive evidence alone (Pullum & Scholz, 2002; Rohde & Plaut, 1999). For some time, it was suggested that infants might actually receive some negative evidence, since parents might occasionally correct their children's mistakes, or fail to understand their severely ungrammatical utterances. As later studies clarified (for summaries, see Pinker, 1984; Marcus, 1993; Guasti, 2002), parents actually do not frequently correct their offsprings, and when they do, they correct factual mistakes, rather than grammatical errors. More importantly, infants have been experimentally shown not to benefit from correction (Marcus, 1993; Guasti, 2002). If they don't know a construction, they will not be able to produce it correctly even when repeating after a model. In sum, it has clearly been established that no negative evidence is reliably and systematically available to infant learners.

Although logically less detrimental to learning, it has also been observed that the input is often fragmentary and 'noisy', containing agrammatisms, hesitations, false starts, reiterations, etc. Inter- and intraindividual variation, such as changes in style, register, geographical dialect and the like, also introduce 'noise' into the input. Moreover, the quantity, quality and contents of the received input also vary from one child to the other, thus some children may get exposed to certain constructions later than others or never at all.

These logical problems are best illustrated through two simple examples.

undescribed native American tongues, which are typologically very different from Indo-European languages. Given the methodological stance of structuralism, they proposed strictly statistical and computational methods to solve the problem (for details, see the discussion of Harris (1955) in section 3.1.1). However, they were well aware of the difficulty, *voire* impossibility of such a bottom-up, inductivist approach (Quine, 1953).

Chomsky (1980) notes that the question in (1b) could be derived from the affirmative in (1a) following an infinite number of rules, e.g. (i) 'Move the first auxiliary to the beginning of the sentence', (ii) 'Move the auxiliary belonging to the subject of the main clause to the beginning of the sentence', (iii) 'Move all auxiliaries to the beginning of the sentence', etc. Even if further questions occur in the input, most of the above possibilities cannot be excluded. Yet, only one of them is adequate to characterize the grammar of English. As shown in the more complex context of an embedded clause (2), only rule (ii) yields a grammatical question ((2c) vs. (2b)-(2d)). If a learner is only exposed to simple questions like (1), which Chomsky assumes to be the case for at least some English learning infants, there is no way to exclude rules (i) and (iii) purely on the basis of the input and in the absence of negative evidence. Therefore, it is impossible to arrive at the correct generalization. Nevertheless, English learning infants, even if never exposed to complex examples (2) providing the crucial piece of evidence, have never been observed to produce (2b), (2d) or any other ungrammatical alternative (Crain & Nakayama, 1987; but see also Ambridge, Rowland, & Pine, in preparation for some evidence to the contrary).

(1) a. *The man is tall.*
b. *Is the man tall?*

(2) a. *The man who is in the garden is tall.*
b. **Is the man who _ in the garden is tall?*
c. *Is the man who is in the garden _ tall?*
d. **Is is the man who _ in the garden _ tall?*

All these problems notwithstanding, all healthy human infants acquire language. What is particularly surprising about this is (i) that they acquire the very language spoken in their environment, converging by and large on the same grammar as the other members of their linguistic community, and (ii) that the grammar they have acquired allows them to go beyond the input they had been exposed to. To go back to the previous examples, even if never exposed to auxiliary extraction from an embedded clause, human infants know that the auxiliary to be fronted is the one that structurally belongs to the subject of the main clause, not the one that comes first. This suggests that they have a priori expectations about what form possible grammatical rules can take. In the given example, what distinguishes rule (ii) from the other two is that it is structure-sensitive, referring to notions such as 'subject', 'main clause' or the dependency between the subject and the auxiliary, whereas rule (i) refers to serial position, rule (iii) to category membership, but not function and structure.

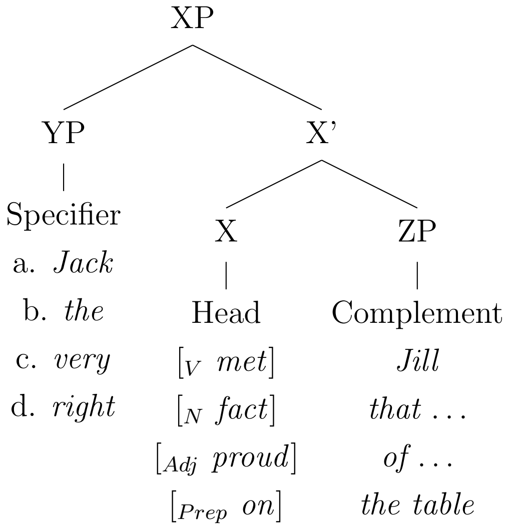
1.1.2 The Principles and Parameters (P&P) theory of language and language acquisition

Since evidence from the input is, by logical necessity, insufficient to induce the correct grammar, Chomsky (1959, 2000) has argued that humans must be innately equipped with constraints or mechanisms that allow them to entertain only the linguistically meaningful rules and select between possible grammars. This innate knowledge, called Universal Grammar (UG), encoded in the biological endowment of our species, consists of Principles, defining linguistic properties that universally characterize all human languages, and Parameters, encoding properties that vary across the languages of the world. Thus Principles represent the commonalities, while Parameters account for the observable surface differences of languages. Consequently, Principles and Parameters together define the logical space in which all natural languages necessarily fall.

As an example of a UG principle, all human languages distinguish between different categories of lexical items, the most general distinction holding between content words carrying lexical meaning (e.g. nouns: *car, dog, peace* etc.; verbs: *eat, hit, think* etc.; adjective: *red, good, incredible* etc.) and functors encoding grammatical relations (e.g. determiners: *a, the, some, these* etc.; pronouns: *I, he, hers, us* etc.; pre- or postpositions: *up, of, on* etc.).

Also, all languages of the world break their sentences up into hierarchically organized syntactic units. The most basic unit is the syntactic phrase (Chomsky, 1970), comprising of a Head, which determines the type and the syntactic behavior of the phase and subcategorizes for a Complement and a Specifier (3).

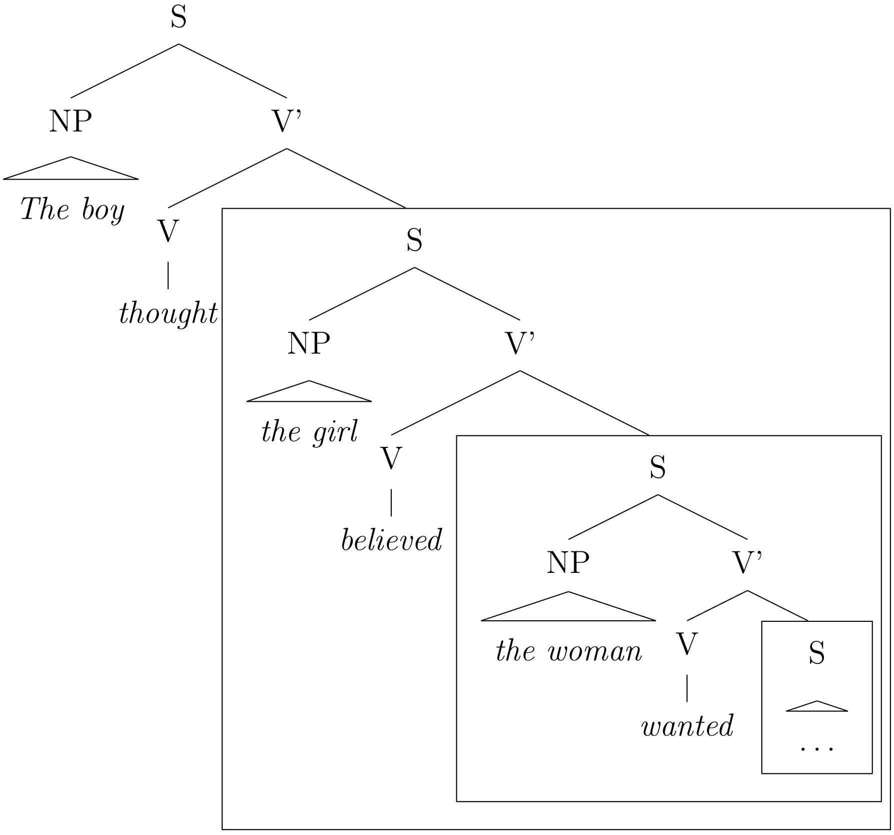
(3)



[a: Verb Phrase (VP); b: Noun Phrase (NP); c: Adjectival Phrase (AP); d: Prepositional Phrase (PP)]

A third fundamental principle of human languages (Chomsky, 1957, 1995) is that the above phrasal schema, known as $\overline{X}$ theory, allows an *infinite* and *recursive* combination of syntactic units, since any phrase can be extended by inserting another phrase, including one of the same type (4), into its Specifier or, more typically, Complement position. These combinatorial properties make human languages an open and discretely infinite generative system with the power to produce an *infinite* number of sentences, already heard or completely novel, using *finite* resources, viz. a finite, though extendable vocabulary, and a finite set of rules. The creativity ensured by the discrete infinity of UG is what allows infants to go beyond the input and produce sentences never before encountered.

(4)



[S: Sentence[4]; NP: Noun Phrase; the triangle indicates unanalyzed subtrees]

Parameters, on the other hand, are best understood as 'switches' that configure certain linguistic constructions. For instance, while the grammatical relations encoded in the $\overline{X}$ schema are universal, the linear order in which the three constituents follow each other varies among languages in systematic ways. This is captured by two word order parameters, the Head-Specifier and the Head-Complement parameters, configuring the order of the respective phrasal constituents. Their different settings give rise to six possible orders, all of which are actually attested in the world's languages (Dryer, 1992). Taking the VP as an example, with the Subject as its Specifier and the Object as its Complement, Table 1.1 exemplifies the six different ba-

[4] The S used in this simplified tree telescopes VP and a number of other phrases (agreement, tense etc.) that, in a full representation, would intervene between S and VP.

sic word orders and provides the percentage of their occurrence among the languages of the world, as reported in two different studies (Mallinson & Blake, 1981; Ruhlen, 1975). Interestingly, not all patterns appear with equal frequency. SOV and SVO account for the vast majority of languages, while O-initial languages are extremely rare. This suggests that while all settings of the parameters are possible, there are preferred values, and the parameters, although independent, are not unrelated.

Word Order	Ruhlen (1975)	Mallinson and Blake (1981)
SOV	51.5%	41%
SVO	35.6%	35%
VSO	10.5%	9%
VOS	2.1%	2%
OVS	0%	1%
OSV	0.2%	1%
Unclassified	–	11%

Table 1.1: Basic word order frequencies in two language samples.

Another example of a major typological parameter is *pro*-drop (Rizzi, 1986). Languages systematically differ in whether they allow pronominal arguments, especially subjects, to be omitted. In English, for instance, the *pro*-drop parameter is set to the negative value. Noun phrase arguments have to be overtly specified even if (i) pronominal and (ii) semantically void, as in the case of expletive subjects (5a)[5], which, since semantically empty, are required for purely grammatical reasons. In Italian, on the other hand, *pro*-drop is set to the positive value, since pronominal subjects can (but need not) be dropped (5c). Crucially, the parameter provides more than just a simple redescription of the data, as it governs the behavior of a series of other linguistic properties, which, on the surface, seem unrelated to the presence or absence of pronominal arguments. Italian, for example, allows postverbal subjects, while English doesn't (6). This difference has also been linked to the different setting of the *pro*-drop parameter in the two languages.

[5] Abbreviations:
3sg: 3rd person singular; prpart: present participle; nom: Nominative case

(5) a. ***It*** *is raining.*

b. *_ piove.*
rain.3SG
'It rains.'

c. (i) *_ sta mangiando.*
AUX.3SG eat.PRPART
'(He) is eating.'

(ii) *Gianni sta mangiando.*
Gianni.NOM AUX.3SG eat.PRPART
'Gianni is eating.'

(6) a. (i) *John arrived.*

(ii) **Arrived John.*

b. (i) *Gianni è arrivato.*
Gianni.NOM AUX arrived
'Gianni (has) arrived.'

(ii) *È arrivato Gianni.*

Currently, there is no consensus about the exact number or the precise form of the parameters. Estimates concerning the number of necessary parameters vary from a dozen to a few hundred, but it is generally desired (and expected for reasons of parsimony and minimality) that their number be small (Longobardi, 2005). Also, while no formal description is widely accepted, it is often assumed that parameters are binary (Chomsky, 1995). Several theorists have also proposed that parameters might be hierarchically organized (Baker, 2001; Longobardi, 2001, 2005). Under this view, certain parameters are more important, more general than others, and their settings determine the values of dependent, lower-level parameters, or whether the low-level parameters apply in a given language at all.

In the P&P theory, acquiring language corresponds to setting the parameters to the values that characterize the target grammar (Wexler, 1998; Chomsky, 2000; Rizzi, 2005). Since parameters are universal and prewired, they need not be learned. What is to be acquired is only the variation between languages, and this, as discussed above, is encoded in the parameters. Therefore, the learner's task is to set the parameters. This can be regarded as yet another case of 'learning by forgetting' (Mehler & Dupoux, 1994), common across several domains in language acquisition, whereby the mature system is reached by eliminating certain options from an initially given superset of choices. Neonates, for instance, are able to distinguish all the phonemes found in the world's languages, but after a few months of experience with the target language, they lose this universal ability and only discriminate native phoneme distinctions (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1983, 1984; Tees & Werker, 1984). Similarly, all parameter settings are initially possible, and infants have to converge on the subset that adequately characterizes their target language. Although not strictly a logical necessity in this framework, parameter setting accounts come with the strong implication that there is a continuity between child and adult grammars (Pinker, 1984; Rizzi, 2005). Moreover, inasmuch as a child grammar differs from the target adult grammar, the discrepancies have to be grammatical options that may not characterize the adult grammar of the target language, but at least the adult grammar of some other language.

Parameters are assumed to be set by certain pieces of information in the input that work as triggers. For instance, the *pro*-drop parameter (Rizzi, 2005) can easily be set to the positive value in Italian given the presence of a large number of sentences with null subjects and rich verbal morphology. It is an open question (Chomsky, 2004) whether parameters come in a neutral setting at the initial state, or whether there is a preferred default value, as seem to be suggested by some typological data and errors in infants' early production. It has been proposed that some major parameters, e.g. word order parameters, are set correctly prior to any syntactically complex production ('Very Early Parameter Setting', Wexler, 1998). Others, mostly related to the optional deletion of material (e.g. *pro*-drop, copula-drop or article drop) are universally set to the positive value initially, giving rise to early production errors and resetting, in languages that instantiate the negative value (Rizzi, 2005). For example, child English is characterized by frequent subject drops in affirmative finite main clauses (e.g. (7) from R. Brown, 1973).[6]

[6] Rizzi (2005) raises an interesting issue with respect to these initially misset parameters.

(7) _ falled in the briefcase. (Eve 1;10)

Parameter setting provides an elegant conceptualization of the problem of language acquisition. However, it does not provide a full solution, since parameters are abstract, discrete and symbolic linguistic 'entities', which need to be linked somehow to the concrete and continuous input. The Head-Complement and Specifier-Head parameters, for example, are useful formal tools to characterize word order. However, the child cannot directly apply them to the input, since words in the input do not come labeled as Head, Complement or Specifier.

This 'linking problem' (Pinker, 1984) has drawn theorists' attention to the information contained in the input, seeking cues that can help infants bridge the gap between abstract linguistic knowledge and concrete linguistic signal.

1.2 The richness of the signal: statistics and bootstrapping

The properties of the signal have passed into the forefront of attention because they constitute a logical boundary condition for language acquisition theory. As discussed above, the poverty of the stimulus and the lack of negative evidence have been central to the claims about universal grammar, providing a 'lower boundary' for acquisition theory and supporting the innateness hypothesis. On the other hand, the information that *is* actually contained in the signal acts as an 'upper boundary': the language faculty need not encode contents that are available in the input. To put it differently, in order to understand how acquisition mechanisms work, it is necessary to know what information they are designed to learn. In Morgan and Demuth's (1996, p. 3) formulation: "Inclusion of appropriately rich representation of

While most accounts that argue for some default parameter value rely on the unmarkedness of the given typological choice, Rizzi claims that default parameter values might be those that facilitate production (e.g. by allowing the dropping of material), thereby favoring the immature performance system of young learners.

the input may entail modifications of theories of grammatical development, particularly with regard to theoretical characterizations of the initial state."

There have emerged two interrelated lines of research investigating the information contained in the signal, based on two learning mechanisms mentioned before, i.e. rule extraction and statistical learning. Bootstrapping theories have been focusing on the linking problem, seeking to identify perceptually available features of the signal that reliably co-occur with some abstract property and might thus serve as a cue to it. The other approach, statistical learning theory, has revived earlier structuralist and information theoretic models, arguing that the probability distributions characteristic of linguistic constructions allow learners to pick up statistically coherent or frequent patterns from the input. Below, I will introduce these two approaches in turn.

1.2.1 Bootstrapping syntax

While the syntactic/structural properties of words and sentences are not overtly manifest in the input, they are often accompanied by other features of the signal that are perceptually available. Nouns and verbs, for instance, are abstract lexical categories. However, in English, nouns often bear stress on the first syllable (record N: /ˈrekɚd/), verbs on the last (record V: /rɪˈkːɔrd/) (Cutler & Carter, 1987; Davis & Kelly, 1997). The stress pattern, then, can act as a cue to the two categories. While this cue is specific to English, in other languages, other regularities of this kind may be present.

In more general terms, bootstrapping theories argue that associations between perceptually available surface cues and abstract, perceptually unavailable structural patterns can be used to solve the linking problem, and set syntactic parameters. As Morgan and Demuth (1996, p. 2) put it: "[T]hese [=bootstrapping] accounts propose that information available in speech may contain clues to certain fundamental syntactic distinctions [...]."

This reasoning assumes that (i) such associations exist, (ii) infants are sensitive to the surface bootstrapping cue, and (iii) they 'know', i.e. it is somehow encoded in their language faculty, that a certain cue bootstraps a certain structural parameter. It is an open question whether such associations are purely accidental correlations or whether there are yet unknown (indirect) causal links between surface cues and morphosyntactic properties they bootstrap. Note, however, that logically, it is more parsimonious to assume that the associations are causal. Otherwise, we need not posit that the language faculty contains arbitrary relations, whose origin would be hard to explain.

Several bootstrapping mechanisms have been proposed, making use of different surface cues as triggers. One approach (e.g. Pinker, 1984) suggests that the relevant cue is of semantic/conceptual nature. By understanding the general meaning of some simple sentences, and by knowing the meaning of some words, the infant can construct syntactic trees, given configurational universals, such as $\overline{X}$ theory, contained in her language faculty. From these trees, the child can derive the syntactic rules of her mother tongue, which in turn, help her parse and understand more complex sentences.

A second approach (e.g. Gleitman & Landau, 1994) claims that the already acquired pieces of syntactic knowledge help bootstrap the rest of syntax. The initial (productive) lexicon of the child contains a large number of nouns. This allows the infant to track the position of nouns within sentences. With this information, infants can learn the type and argument structure of verbs. In English, for instance, intransitive verbs have one noun (phrase) preceding them, transitive action verbs have one noun (phrase) preceding and one following them, mental verbs have one noun (phrase) preceding them and a clause following them, and so forth. Thus upon encountering a sentence containing an initial NP and a final NP with a verb between them, the verb can be categorized as transitive.

A third approach suggests that structural properties are signaled by their acoustic/phonological correlates (Morgan & Demuth, 1996; Nespor, Guasti, & Christophe, 1996; Mehler, Sebastian Gallés, & Nespor, 2004; Nespor et al., under review). This approach, unlike the others, assumes no prior linguistic knowledge on the part of the learner, and thus may be insightful in explaining the earliest acquisitions. One proposal in this line of research has focused on the acquisition of word order, especially the setting of the

Head-Complement parameter. Nespor et al. (1996) argue that the position of prosodic prominence in phonological phrases correlates with word order, and can thus provide a perceptually available surface cue to it. In Turkish, for example, which is an OV language, the prominence is left-most, i.e. phrase-initial (e.g. ***kilim*** *için* kilim for 'for the kilim')[7], while in the VO language French, prominence is right-most, i.e. phrase-final (e.g. *pour châque* ***morale*** 'for each ethic'). Moreover, this cue is perceptually detectable, since phrase-initial prominence is proposed to be cross-linguistically realized as increased pitch and intensity, while phrase-final prominence is mainly marked by increased duration (Nespor et al., under review). This is true, even when the same phrase type exhibits both word orders. In Dutch, for instance, if a prepositional phrase is pronounced with its canonical preposition-noun order (*op de* ***trap*** 'up the stairs'), prominence is realized by lengthening the noun. If, on the contrary, the phrase has a non-canonical noun-preposition order (*de* ***trap*** *op* 'the stairs up'), motivated by certain pragmatic contexts, prominence is implemented as higher pitch and intensity on the noun (which is now to the left). Nespor et al. (1996) have also shown that infants are sensitive to this prosodic cue, distinguishing French and Turkish stimuli solely on the basis of prosodic prominence (as other phonological cues were removed).

Building on similar empirical observations, another interesting proposal (Mehler et al., 2004) has been motivated by the well established typological fact that there is a very general universal correlation (or maybe causal relation) between a number of phonological, morphological and syntactic properties, such as the syllabic repertoire, the ratio of vowels and consonants, the place and realization of prosodic prominence, the morphological type, and the basic word order of a language (Fenk-Oczlon & Fenk, 2005). Languages like Japanese or Turkish, having relatively simple syllabic structure, high vocalic ratios, and leftward prominence, tend to have agglutinating morphology and OV basic word order. Languages like Dutch or Polish, which have

[7] In-text examples follow the formatting convention of the numbered, indented examples, i.e. the original example is set in italics, it is followed by the morphemic glosses, then the English translation is given between single quotes. No morphemic glosses are given, if the order and function of the morphemes in the original are equivalent with those in the English translation.

very complex syllables, low vocalic ratios and rightward prominence, tend to be non-agglutinating (typically inflecting) and have VO basic word order. Ramus and Mehler (1999) and Mehler et al. (2004), building on original observations by Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, and Mehler (1988), have quantified this observation by measuring syllabic complexity (ΔC) and vocalic ratios (%V). ΔC is the variability of the amount of time spent on consonants in the speech stream. If a language has simple syllabic structure, e.g. Japanese, which basically only allows (C)V(N)[8] syllables (e.g. *Hon-da, To-ky-o, i-ke-ba-na*), ΔC will be low since the length of consonant clusters varies little, i.e. one C or nil. In languages that have complex syllabic structure, like English with its (C)(C)(C)V(C)(C)(C) (e.g. *eye* /aɪ/ vs. *strengths* /streŋθ/), consonant clusters vary greatly in length, and produce high ΔC. %V is the amount of time spent on vowels in the speech stream, i.e. the proportion of vocalic time relative to the total length of the stimuli. If consonant clusters are short in a language, e.g. Japanese, %V will be high. If consonant clusters tend to be long, e.g. Croatian *prst* /prst/ 'finger' with no vowel at all, %V will be low. Mehler et al. (2004) plotted a number of typologically, genealogically and geographically different languages in the two dimensional space defined by ΔC and %V, expanding earlier work by Ramus and Mehler (1999) and Ramus (2002). They have found that languages clustering together in this space do indeed have similar prosodic prominence, morphological type and word order. While the correlation between the cues seems to be robust, there exists only indirect experimental evidence that infants use these complex cues. Ramus and Mehler (1999) and Ramus (2002) have shown that newborns are able to distinguish languages on the basis of their different %V values. However, there is no demonstration that infants might link this cue to morphosyntactic properties. Yet, the proposal relies on robust typological evidence, and if true, it has the potential to account for the early acquisition of most major typological differences among languages.

[8] Parentheses indicate optional phonemes. N stands for any of the nasals /n/, /m/ and /ŋ/. Additionally, Japanese also allows (C)V(C*) syllables, where C* is the first half of a geminate whose second half is the C of the next syllable, such as in the city name *Sap-poro*. However, geminates are not frequent and only a restricted number of consonants, i.e. only obstruents, are allowed to geminate in Japanese.

As is apparent from the above discussion, parameter setting and bootstrapping are specific formalisms to capture the mechanisms of rule extraction and generalization. Following Chomsky's (1957, 1959) early contributions, the definitions of rule extraction, generalization and rule-governed behavior have been heatedly debated in the philosophical, epistemological and psychological literature. While fully acknowledging the complexity of the issue, I will assume the following definition of rule extraction and generalization for the purposes of the present work. Rule extraction (or generalization) is a learning mechanism that posits open-ended representations, i.e. representations containing a variable, over a set of data. A variable is a placeholder or open slot into which *all* items of a category can be inserted. In language acquisition, rule extraction or generalization typically happens on the basis of a rather limited set of data, i.e. sparse input (Newport, 1990; Endress & Bonatti, 2006).[9] Yet, crucially, it allows to go beyond the input and make correct inferences about novel instances of the category represented by the variable.

In sum, bootstrapping hypotheses provide explanations about how abstract structure might be learnt in the absence of explicit evidence. They build upon the correlations that exist between the perceptible and the underlying properties of language, ultimately relying on the assumption (Chomsky, 1959) that the architecture of natural language is, in part, shaped by learnability constraints.

1.2.2 Statistical learning: segmentation and word learning

Since the beginnings of American structuralist linguistics (Harris, 1951, 1955) and information theory (Shannon, 1948), it has been recognized that the

[9] Strictly speaking, the definition of rule extraction, unlike that of generalization, does not need to include the limited nature of the data set. It is possible that a rule is extracted after having met all the instances of a given category. However, given that content word categories, e.g. nouns, verbs etc., are open classes, rule extraction typically implies a limited set of data. Therefore, I do not draw a sharp distinction between rule extraction and generalization here.

linguistic code is statistically informative: certain units are more likely to occur than others, and this is modulated by the context.

After the seminal work of J. Hayes and Clark (1970), who showed that adults are able to segment a continuous stream made up of square wave analogues of speech upon mere exposure to it using the statistical information in the stream, Saffran, Aslin, and Newport (1996) investigated whether this mechanism is also available to infants. They created an artificial language consisting of four trisyllabic nonsense words (e.g. “bidaku”, “padoti” etc.), which were repeated in random order. The language 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 transition probability[10] (TP) of 1, while syllables spanning word boundaries had a TP of 0.33. Two minutes of exposure to the artificial language was enough for 8-month-old infants to discriminate between ‘words’ and trisyllabic ‘part-words’, which were obtained by concatenating the last syllable of a word and the first two syllables of another word, i.e. they were chunks that actually occurred in the language, but had the wrong statistical structure, containing a drop in TPs word-internally.

These findings have given rise to a rich body of research, investigating different properties of statistical learning (Bonatti, Peña, Nespor, & Mehler, 2005; Fiser & Aslin, 2002; Hauser et al., 2001; Newport & Aslin, 2004; Newport, Hauser, Spaepen, & Aslin, 2004; Peña, Bonatti, Nespor, & Mehler, 2002; Toro & Trobalon, 2005; Toro, Bonatti, Nespor, & Mehler, in press; Thiessen & Saffran, 2003; Shukla, Nespor, & Mehler, 2007). Statistical learning has been shown to be a robust, domain-general, age-independent and not specifically human ability. It operates not only over linguistic stimuli, but also tones (Nonaka, Kudo, Okanoya, & Mizuno, 2006) and visual input (Fiser & Aslin, 2002). It is performed by newborns (Nonaka et al., 2006; Teinonen, 2007), infants at 8 and 13 months (Saffran, Aslin, & New-

[10] Forward transition probability is the conditional probability of a unit Y to appear (immediately) after unit X. Formally, TP(X→Y)=F(X)/F(XY), where F(X) is the frequency of X, F(XY) is the frequency of XY. For more on conditional probabilities, see Chapter 3.

port, 1996; Marchetto & Bonatti, in preparation), and adults (Peña et al., 2002). Moreover, non-human species, such as tamarin monkeys (Hauser et al., 2001) and rats (Toro & Trobalon, 2005) are also able to learn statistical information.

A set of studies have focused specifically on the relevance of TP computations for language acquisition. Inspired by the fact that both morphology and syntax make use of constructions with distant dependencies, Peña et al. (2002) and Newport and Aslin (2004) asked the question whether transition probabilities between non-adjacent items can be learnt. The first group of authors found that adults readily segmented out trisyllabic words from an artificial language when they were defined by high TPs between the first and the last syllables (A X C). However, subjects failed to generalize the pattern unless (subliminal) segmentation cues were inserted into the stream to facilitate the original segmentation task. The second group of authors, in contrast, found that adult subjects were poor at segmenting when the non-adjacent regularity applied between syllables, but they were successful when it applied between phonemes (consonants and vowels, invariably).

These results lead to a second issue that is important for language acquisition: the units or representations used for statistical computations. While Newport and Aslin (2004), as mentioned above, found good segmentation for both vowels and consonants, Bonatti et al. (2005) observed that adults readily segment over non-adjacent consonants, but not over non-adjacent vowels. It is not yet clear why the two groups have found different results, but one factor might be the structure of the familiarization stream. The one that Newport and Aslin (2004) used allowed immediate repetitions of the same word frame, whereas Bonatti et al.'s (2005) stream had no immediate repetitions.

Further investigating the question of representational units, Toro et al. (in press) devised a series of artificial grammar experiments to show that consonants and vowels serve as preferential input to different kinds of learning mechanisms. They found that participants performed well when their task was to do statistical computations over consonants and rule-learning over vowels (the rule to be learnt was a repetition-based generalization). But their performance dropped to chance in the opposite case. Taken together, these studies indicate that not all linguistic representations are equally suitable for statistical learning. Consonants seem to be the primary target, while vowels are preferentially recruited for rule learning. These latter can be used for statistical computations only under special conditions, such as the informationally highly redundant stream used by Newport and Aslin (2004). These findings converge with certain observations in theoretical linguistics (Nespor, Peña, & Mehler, 2003) claiming that consonants and vowels have different linguistic functions. Consonants are believed to be responsible for encoding the lexicon, e.g. consonantal stems carry the semantic contents of lexical items in Semitic languages, whereas vowels are claimed to signal morphological form and syntactic function, e.g. Ablaut phenomena in Germanic languages, *s**i**ng, s**a**ng, s**u**ng.*

A third issue that has been raised is how statistical computations interact with other mechanisms that signal boundaries in the input, e.g. word stress and prosody. Thiessen and Saffran (2003) have shown that 9-month-old, but not 7-month-old English-learning infants use syllable stress as a cue for segmentation, and not statistics. Since the boundaries of prosodic units coincide with word boundaries, and thus provide reliable and perceptually detectable cues to them, Shukla et al. (2007) have investigated how prosody interacts with statistics. Using an artificial speech stream that had prosodic contours overlaid on it, the authors found that adults compute TPs for all syllable pairs, but reject those that span prosodic boundaries, even if they have high TPs between them. In other words, prosody acts as a filter over the output of TP computations.

Another way in which statistical information is believed to be useful during language acquisition (Tomasello, 2000; Rowland, 2007; Ambridge et al., in preparation) is by providing ready-made constructs of frequently co-occurring elements. Under this constructivist view, language acquisition, at least initially, doesn't proceed through the extraction of abstract rules or the setting of formal parameters. Rather, in the beginning concrete chunks of the input are memorized. Then, in a second step, similarities (e.g. the same words) are discovered between the chunks, giving rise to semi-abstract constructions, with a variable element inside, although with a semantically or pragmatically constrained range. For instance, the variable in the frame *Can the _ go?* does not initially extend to all nouns, but maybe only to members of the family or only to animals etc. Later, syntactic rules are assumed to emerge by generalizing even further over these semi-abstract constructions (Tomasello, 2000).

In this approach, word order is also assumed (Tomasello, 2000; Chang, Lieven, & Tomasello, under review) to be learned from frequently encountered examples in the input. Since young learners are sensitive to co-occurrence statistics in the language they hear, their early competence contains semi-abstract constructions derived from this statistical information. For instance, from frequent occurrences of *Can you see...? Can you go...? Can you eat ...?*, the infant might construct the semi-general frame *Can you X?*, where X is a placeholder for possible substitutions, in this case, for certain verbs. Thus, this view claims that young learners have no general and fully abstract representations of word order. Rather, their knowledge is linked to specific lexical items or frames. This view, then, implies that learning word order proceeds together with building the lexicon.

In the light of the above findings, I will assume the following operational definition of statistical learning in language acquisition (not denying, of course, that other definitions are possible). Statistical learning is a mechanism that collects information about the frequency and probability distributions of items found in the input, and allows this information to be used in certain tasks, such as word learning. In its simplest form, statistical learning collects information over surface instances. Therefore, such learning is typically item-based, or concerns, at most, a limited set of items. (Statistical learning is, of course, also possible over abstract categories of items, but in that case, it requires an abstraction/generalization mechanism prior to or in conjunction with its application.) It is, by its very nature, heuristic/probabilistic, i.e. it leads to correct results in most, but not necessarily in all cases. Given the two previous properties, statistical learning does not give rise to overarching generalizations; its capacity to apply to novel items not encountered before and to go beyond the original input is limited. Another hallmark of this learning mechanism is that it requires a certain amount of exposure. In most cases, sparse data is detrimental to statistical learning, because small samples might provide unrepresentative distributions.

As shown above, exploring the information contained in the signal have uncovered two learning mechanisms, bootstrapping and statistical learning, available to young learners during early language acquisition.

1.3 Perceptual constraints on learning

There is at least one more mechanism that probably plays an important, yet unexplored role in language acquisition. Since the advent of Gestalt psychology, it has been well known that the perceptual system processes certain feature configurations more efficiently and more automatically than others. Yet, the effects of such auditory Gestalts on the acquisition procedure have remained little explored.

Recently, Endress et al. (2005) have proposed that certain results in artificial grammar learning, originally attributed to symbolic rule extraction, are better explained in terms of such perceptual Gestalts or 'perceptual primitives'. Specifically, Marcus et al. (1999) argued that 7-month-old infants distinguish artificial grammars containing repetitions at different positions, e.g. ABB ("wo fe fe"), AAB ("wo wo fe") and ABA ("wo fe wo"), by extracting an abstract identity relation between the syllables and encoding it in a symbolic way. Endress et al. (2005) claimed that infants may distinguish these grammars not because they encode the underlying structure, but because the items contain perceptually salient repetitions at the edge positions. Indeed, in Endress et al.'s (2005) artificial grammar learning experiments, which used longer items, adults were able to tell apart grammatical items from ungrammatical ones only when the repetitions were at the edges, e.g. ABCDEFF, but not when they were in the middle of the item, e.g. ABCDDEF. In another set of experiments, Endress et al. (in press) have also shown that repetitions themselves are special, since adults readily learn tone sequences that contain a repetition, e.g. ABB: low-high-high, but not sequences that contain a similarly regular, but ordinal pattern, e.g. ABC:

low-high-middle, although the inequality relations A $<$ B and B $>$ C could have been easily encoded symbolically.

I operationally define perceptual primitives as configurations of objects or features in the input that the perceptual system detects in an automatic fashion due to the functioning of its neural apparatus. In other words, perceptual primitives are the preferred input configurations of a given sensory system.

1.4 The aim of the thesis

In the light of the above, the present thesis explores language acquisition from two complementary directions: from the input and from the learning mechanisms.

I investigate the properties of the input, focusing on the type of statistical information that is available to learners of typologically different languages. As described above, there is a large body of evidence showing that learners readily pick up (certain kinds of) statistical information from the input. It is also clear that the linguistic signal is rich in statistical information. However, it is not well known whether this information is of the relevant kind and whether it reliably signals morphological boundaries. Therefore, I examine these issues calculating conditional probabilities and frequencies for corpora of infant-directed speech in three typologically, genealogically and geographically different language, Japanese, Hungarian and Italian. The ultimate question is whether statistics in the input universally reflects the morphosyntactic properties of languages.

I also investigate the three learning mechanisms introduced above. I seek to answer two main questions. First, what mechanisms play a role at the very beginning of the acquisition procedure, when infants are about to discover that the linguistic input contains certain regularities? This question brings us closer to understanding the nature of the initial state. I address this issue in a series of optical brain imaging studies with neonates, showing that the neonate brain is already capable of distinguishing structured from unstructured input, readily detecting certain perceptual primitives such as adjacent repetitions, over which it can build abstract representations. The second relates to the mechanisms that play a role in acquiring language-specific knowledge about the native language. Specifically, I will focus on how basic word order might initially be acquired. This question concerns the state of the language faculty when populations acquiring different native languages start to diverge. In a series of artificial grammar learning experiments with 8-month-old Italian and Japanese infants, as well as Basque, Japanese, Hungarian, French and Italian adults, I find evidence for a frequency-based bootstrapping mechanism potentially cuing basic word order.

As a synthesis of exploring the three learning mechanisms at two key stages of linguistic development, I aim at developing an integrative model of language acquisition.

As mentioned above, the thesis focuses on the earliest acquisitions of grammar, in particular word order and morphological type. These linguistic properties correspond to the most fundamental and most general features of languages, and constitute the major sources of variability among them. Consequently, infants need to learn them from the exposure they receive. Moreover, these properties are prerequisites for the acquisition of more subtle linguistic features. Yet, they are highly abstract and manifest themselves in many different surface forms, so they are not directly extractable from the input. All these properties render the acquisition of word order and morphological type key issues in language development.

To sum up, and to anticipate the main points, the thesis puts forth the following hypotheses:

1. Humans are born equipped with auditory computational primitives that allow them to process and learn certain structural aspects of auditory stimuli immediately at birth. Such primitives can pave the way for other perceptual and symbolic computations that play a role later during language acquisition.

2. The input that young learners receive is rich in—statistical, prosodic etc.—information that correlates with and potentially bootstraps structural categories and regularities.

3. Infants are able to use this information to learn about structure, e.g. word order, independently of and prior to the development of the lexicon.

4. During language acquisition, the genetically endowed abstract linguistic knowledge develops to match the target grammar by relying on information contained in the input and representations sanctioned by the perceptual system (in addition to other, mostly maturational processes).

The thesis is organized as follows. In Chapter 2, aiming at exploring certain aspects of the initial state of the language acquisition faculty, I describe three experiments carried out with neonates using optical topography, which show that the language faculty is equipped with perceptual primitive mechanisms and the ability to generalize these patterns into more abstract representations right from the initial state. In Chapter 3, I examine the linguistic input in a series of experiments on infant-directed Japanese, Hungarian and Italian corpora. A first set of experiments investigates the distribution of different conditional probability measures (forward and backward transition probabilities, and mutual information) and evaluates whether they provide sufficient information for segmentation. In a second set of experiments, I assess frequency as a cue to the categorization of functors and content words, and to basic word order. In Chapter 4, I show that learners are sensitive to this frequency information and might use it to bootstrap grammar. I present two sets of artificial grammar learning experiments, one with adults, the other with infants, showing that learners have an underlying abstract representation of the basic word order of their native language. Building on the findings of the corpus experiments, I propose that such a representation, especially in infants, might be cued by the frequency distribution and the relative order of functors and content words. In Chapter 5, I discuss and synthesize the empirical results, proposing an integrative view of language acquisition. Finally, in Chapter 6, I discuss the broader implications of the findings and the model in the perspective of a general theory of cognitive development.

Chapter 2

What is in the initial state? Perceptual primitives and generalizations: an optical topography study with neonates

What is the genetically endowed toolkit of the language learner? Are the three learning mechanisms involved in language acquisition, namely statistical learning, abstract generalizations and perceptual primitives, already present in the initial state of the language faculty? This Chapter sets out to answer these questions through a series of brain imaging experiments investigating the linguistic abilities of newborn babies.

As mentioned in Chapter 1, newborns have been shown to segment streams of pure tones (Nonaka et al., 2006) and naturalistic syllables (Teinonen, 2007) on the basis of statistical information, as measured by their electrophysiological brain responses in experiments similar to the original (Saffran, Aslin, & Newport, 1996). It is also known that hearing newborns are tuned to the phonological and melodic aspects of spoken language. They are able to distinguish all the phonemes found in the world's languages (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Werker & Tees, 1983; Tees & Werker, 1984) or discriminate unknown languages on the basis of their rhythmic characteristics (Bertoncini et al., 1988; Ramus & Mehler, 1999; Ramus, 2002; Nazzi & Ramus, 2003).

Much scarcer is the evidence about newborns' generalization abilities and the perceptual constraints that their auditory system honors. Therefore, the experiments reported in this Chapter test the presence of these two abilities. In Chapter 1, I have introduced the debate about whether learning ABB and ABA type repetition-based artificial grammars is attributable to symbolic rule-learning (Marcus et al., 1999) or perceptual primitives, such as repetitions at edges (Endress et al., 2005). To tease the two mechanisms apart in newborns, I have designed a series of experiments, in which newborns' brain reactions to different repetition grammars (ABB: adjacent repetitions; ABA: non-adjacent repetitions; A_A: representationally adjacent, temporally non-adjaceent repetition) were compared to random control grammars (ABC in the case of ABB and ABA, and A_C in the case of A_A). Since the participants were exposed to the grammars over a period of more than 20 minutes, both perceptual biases, present from the beginning, and generalization or rule learning, building up over the time course of the experiment, could be tested.

To measure brain responses, I used optical topography (OT)[1], also known as near-infrared spectroscopy (NIRS) (Villringer & Chance, 1997; Meek, 2002; Peña et al., 2003). This technique works somewhat similarly to functional magnetic resonance imaging (fMRI) inasmuch as it also measures neural activity through associated metabolic processes, such as blood flow and blood oxygenation. More specifically, OT measures the concentration changes of oxygenated and deoxygenated hemoglobin (oxyHb & deoxyHb, respectively) through their different absorptions (Figure 2.1) of near-infrared light projected onto participants' heads via optical fibers (Figure 2.2).

This technique is particularly suitable for testing newborns and very young infants, whose skull is thinner and who have much less hair than adults, allowing near-infrared light to penetrated deeper into the cortex (up

[1] For abbreviations, see p. 1.

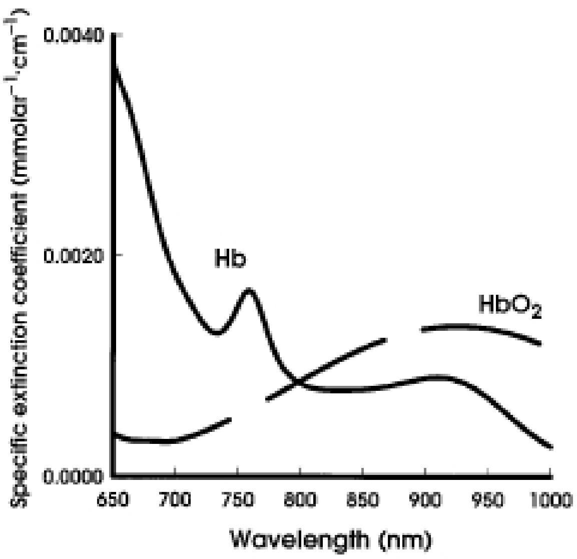


Figure 2.1: The absorption spectrum of oxyHb and deoxyHb in the near-infrared range. Image adapted from Meek (2002).

to about 1.5 cm as opposed to about 0.3–0.5 cm in adults; cf. Figure 2.2). In addition, this technique is fully non-invasive. Unlike fMRI, it is completely silent, and unlike EEG, it does not require the use of carrier substances, such as gels or liquids.

2.1 Experiment 1: Adjacent repetitions (ABB)

This experiment compares a test grammar based on adjacent repetitions (ABB: /talulu/, /penana/, /biʃɔʃɔ/ etc.) to a random control (ABC: /talupi/, /penaku/, /biʃɔge/ etc.), matched to the former in all of its non-structural properties. This first experiment has a double purpose: to establish whether newborns' brains are able to recognize structure in the input at all, and to clarify whether perceptual primitives and generalizations are present in the initial state.

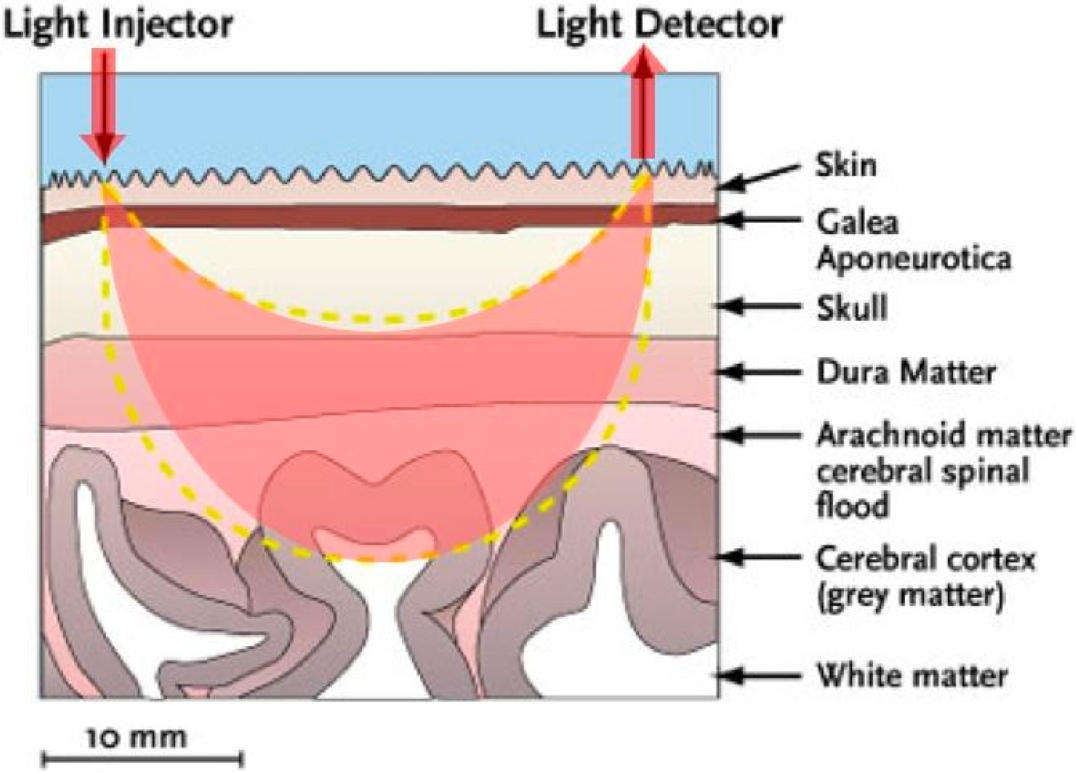


Figure 2.2: The path of the near-infrared light through the human head. Part of the light emitted by the light source over the scalp is scattered. Another part is absorbed by the tissues, in particular by oxygenated and deoxygenated hemoglobin and other chromophors. The remainder, refracted by the tissues, travels over a banana-shaped path and exits the scalp, at which point it is measured by a detector. The model represents an adult head. Original image downloaded from `http://www.the-scientist.com/article/display/15220/`.

2.1.1 Material

Two languages were created, an ABB repetition grammar and an ABC random or control grammar. Both generated trisyllabic 'sentences', but while in the ABB grammar, the second and the third syllables were identical, in the ABC grammar, all syllables were different. Therefore, a regularity, i.e. repetition of the second and third syllables, could be observed in the ABB, but not in the ABC grammar. Both languages used the same syllabic repertoire, containing 20 consonant-vowel (CV) syllables (see Table 2.1), made up of 12 consonants and 5 vowels. The syllables were chosen so that they can be organized into syllable pairs. A syllable pair was defined as two syllables containing the same C, but a different V (e.g. /ba/, /bi/), or at least Cs from the same class (e.g. liquid), and a different V (e.g. /mu/, /na/).

A	**B**
/bi/	/ba/
/du/	/ge/
/pe/	/pi/
/ta/	/tɔ/
/kɔ/	/ku/
/lɔ/	/lu/
/mu/	/na/
/fi/	/fe/
/ʃa/	/ʃɔ/
/ze/	/zi/

Table 2.1: The syllable repertoire used in Experiments 1–3. Syllables are organized into pairs of A and B syllables.

The languages were presented in blocks of 10 sentences (Figure 2.3), 14 blocks per language. The full material was built up from the syllabic repertoire in the following manner. For the ABB language, half of the syllables (see Table 2.1) were designated A syllables, the other half B syllables. The two categories were established such that one member of a syllable pair was assigned to category A, the other to category B. In half of the blocks (the ABB blocks), the syllables of category A were used as the initial unrepeated syllable, and members of category B as the repeated second and third syllable, and inversely in the other half of the blocks (the BAA blocks). Thus each syllable appeared in each sentential position with equal frequency. In addition, each block used different pairings of the A and the B syllables. If, for instance, one block contained the sentence /biʃɔʃɔ/, all others contained combinations of /bi/ with the other members of category B (e.g. /bigege/, /bitɔtɔ/ etc.). In both the ABB and the BAA blocks, two constraints were observed when pairing up A and B syllables within a sentence in order to maximize discriminability: (i) they couldn't contain the same V, (ii) nor could they come from the same syllable pair. This resulted in at least 7 possible sentence combinations for each initial syllable, yielding 7 ABB and 7 BAA blocks. In other words, given (i) the repetition grammar, (ii) the syllabic repertoire and (iii) the constraints on pairing, the 14 blocks exhausted all possible combinations without requiring sentences to be repeated more than once.

The ABC sentences were derived from the ABB sentences by replacing the repeated third syllables of sentences with each other within a block (e.g. /talulu/ and /zepipi/ yielded /talupi/ etc.). Thus 14 blocks of the ABC grammar were obtained. Once again, care was taken to avoid repetitions of identical vowels or consonants within sentences. Also, all syllables appeared in each sentential position with equal frequency.

All sentences were synthesized using the fr4 French female voice of the MBROLA diphone database. All syllables were 270 msec long (C: 120 msec, V: 150 msec), and had a monotonous pitch of 200Hz.

Creating the two grammars in the above manner ensured that they were matched for all properties, except for their structure. More specifically, the two grammars were identical for (i) the overall frequency of all syllables, (ii) the frequency of each syllable in each sentential position, and (iii) for all phonological and prosodic characteristics. Additionally, the distribution of transitional probabilities (Saffran, Aslin, & Newport, 1996) was also equated in the two languages by keeping the TPs as high between certain designated BC syllables as they were between the repeated BB syllables.

Within a block, sentences were separated by pauses of varying length

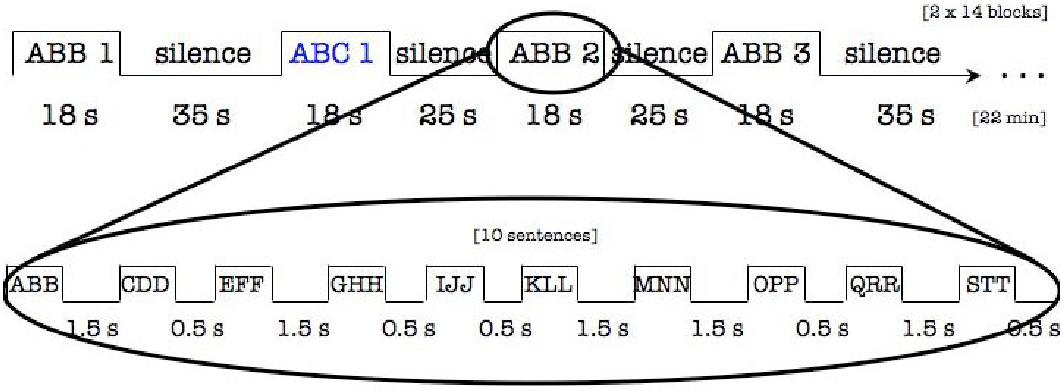


Figure 2.3: The design of Experiments 1–3. The upper boxcar shows the time course of the whole experiment, i.e. the sequence of blocks. The lower boxcar illustrates the sequence of sentences within a block.

(0.5-1.5 sec), yielding blocks of about 18 sec. Blocks were also spaced at time intervals of varying duration (25-35 sec) (see Figure 2.3). This was done in order to avoid phase-locking the brain signal to the stimuli or inducing phase-related brain activity.

The 28 blocks, 14 per condition, were presented in an interleaved fashion, in such a way that at most two consecutive blocks were of the same type. In addition, the order of the blocks was pseudo-randomized and counterbalanced across subjects.

2.1.2 Subjects

Twenty-two healthy, full-term neonates (10 males, 12 females; mean age 3.14 days, range 1-6 days) born to Italian-speaking families participated in the experiment. All infants had Apgar scores ≥ 8 one and five minutes after birth. Parents gave informed consent prior to the experiment. The Ethics Committee of the Azienda Ospedaliero-Universitaria di Udine, where the experiments took place, granted permission.

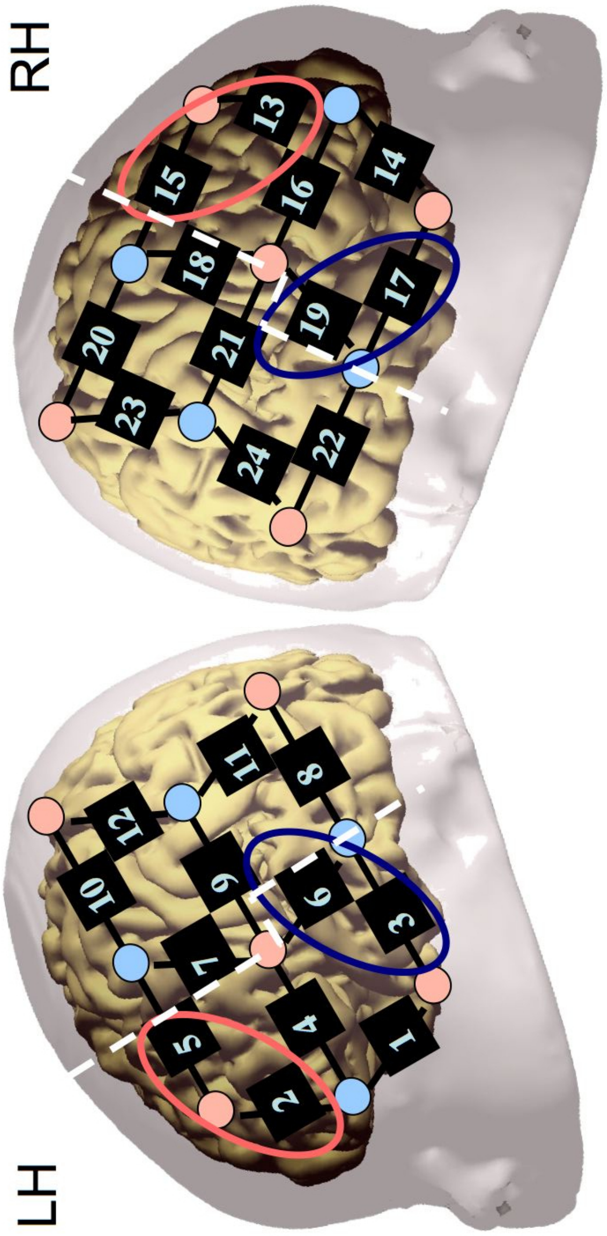
2.1.3 Procedure

Infants were tested in a dimly lit sound-attenuated booth in their hospital environment, lying in their cribs throughout the 22-25 minute-long testing session, assisted by a nurse and an experimenter. Parents could choose to attend the session or not. Babies were tested while in a state of quiet rest or sleep.

Sound stimuli were administered through two loudspeakers positioned at a distance of about 1 meter from the babies' head, at an angle of 30°, and elevated to the same height as the infants' crib. The stimuli were played and the Hitachi ETG-4000 OT machine was operated by a Macintosh PowerPC G5 experimental computer. Both the OT machine and the computer were placed outside the experimental booth. Infants were video-taped during the experiment.

The Hitachi ETG-4000 OT machine used for the experiment had 24 channels (12 per hemisphere), with a source–detector separation of 3 cm. Two continuous light sources using 695nm and 830nm wavelengths with an intensity of 0.30–0.35mW measured light absorption. The silicon probes containing the optical fibers were positioned as indicated in Figure 2.4, using the ears and the vertex as landmarks. This placement allowed to maximize the likelihood of recording from perisylvian and frontal areas.

Figure 2.4 *(following page)*: The placement of the probes on newborns' heads used in all the newborn NIRS Experiments 1–3. Surface landmarks, such as the vertex and the ears, were used to position the probes. LH: left hemisphere. RH: right hemisphere. Red dots indicate light sources, blue dots indicate detectors. Channels are the numbered virtual measurements points between each source–detector pair. Dashed lines separate anterior and posterior regions of interest in both hemispheres. Blue ovals enclose channels assigned to temporal areas of interest in both hemispheres. Red ovals enclose channels assigned to frontal areas of interest in both hemispheres. Infant head and brain model courtesy of Ghislaine Dehaene-Lambertz.


RH
13
14
15
16
18
17
19
20
21
23
24
22
LH
11
8
12
9
10
6
3
7
4
5
1
2

2.1.4 Data Analysis and Statistics

OxyHb and deoxyHb entered into the data analysis. Their concentrations were calculated from the absorption of light recorded by the OT machine.

To eliminate high frequency noises (heartbeat etc.) and overall trends due to systemic changes (blood pressure etc.), the data was band pass filtered between 0.01 and 0.7 Hz. Movement artefacts, defined as concentration changes larger than 0.1 mmol·mm over 0.2 msec, i.e. 2 samples, were removed by rejecting block-channel pairs where artefacts occurred. For the non-rejected blocks, a baseline was linearly fitted between the means of the 5 sec preceding the onset of the block and the 5 sec starting 18 sec after the onset of the block.

To evaluate overall responses to the two grammars, the mean concentration change in oxy- and deoxyHb during the time span of 18 sec starting from the beginning of blocks (corresponding to the time of stimulation) was calculated for each participant over all 14 blocks in each condition, at all channels for which at least two blocks contributed data, i.e. were not rejected. Typically, between 7–14 blocks, i.e. between 50%–100% of the data, were accepted and entered into averaging per channel in each participant. Exceptionally, in one participant, the most posterior channels (11,12 in the LH and 23, 24 in the RH) provided less data (about 4-7 blocks), because excessive head movement caused contact between the probes and the crib/head support. Channels were then further averaged according to their position into left hemisphere (LH) and right hemisphere (RH) channels, as well as into two regions of interest (ROI), anterior and posterior (Figures 2.4), to obtain the predicted locations of differential activation between the two conditions. This measure was used as the dependent variable in a repeated measures analysis of variance (ANOVA) with factors Grammar (ABB/ABC) × Hemoglobin (oxy/deoxy) × Hemisphere (LH/RH) × Area (anterior/posterior).

More specific areas of interests (AOIs) were also defined in order to better evaluate language processing. Auditory processing takes place in the temporal perisylvian areas of the brain (in adults, Friederici, 2002, in infants, Dehaene-Lambertz et al., 2006). Therefore channels 3 and 6 of the LH and channels 17 and 19 of the RH were defined as temporal AOIs. Structural computations are believed to occur in the frontal regions (in adults, Friederici, 2002, in infants, Dehaene-Lambertz et al., 2006). Therefore, channels 2 and 5 of the LH and 13 and 15 of the RH were defined as frontal AOIs. Since only surface landmarks were used to position the probes, the choice of these specific areas is somewhat arbitrary. In addition, anatomic variability across individual participants could not be assessed either. However, channels were chosen with an attempt to cover the relevant AOIs the most uniformly and reliably possible across all babies. A repeated measures ANOVA with factors Grammar (ABB/ABC), Hemisphere (LH/RH) and AOI (frontal/temporal) was conducted to evaluate fine-grained language processing.

To assess the time course of the response and thus evaluate ‘learning’ during the course of the experiment, the means of the first 4 and the last 4 blocks were computed (beginning and end of the experiment, respectively) for each neonate in both conditions for the oxyHb response in the left anterior ROI, which contains most language processing areas, and were entered into a repeated measures ANOVA with factors Grammar (ABB/ABC) × Time (beginning/end).

2.1.5 Results

Overall analysis

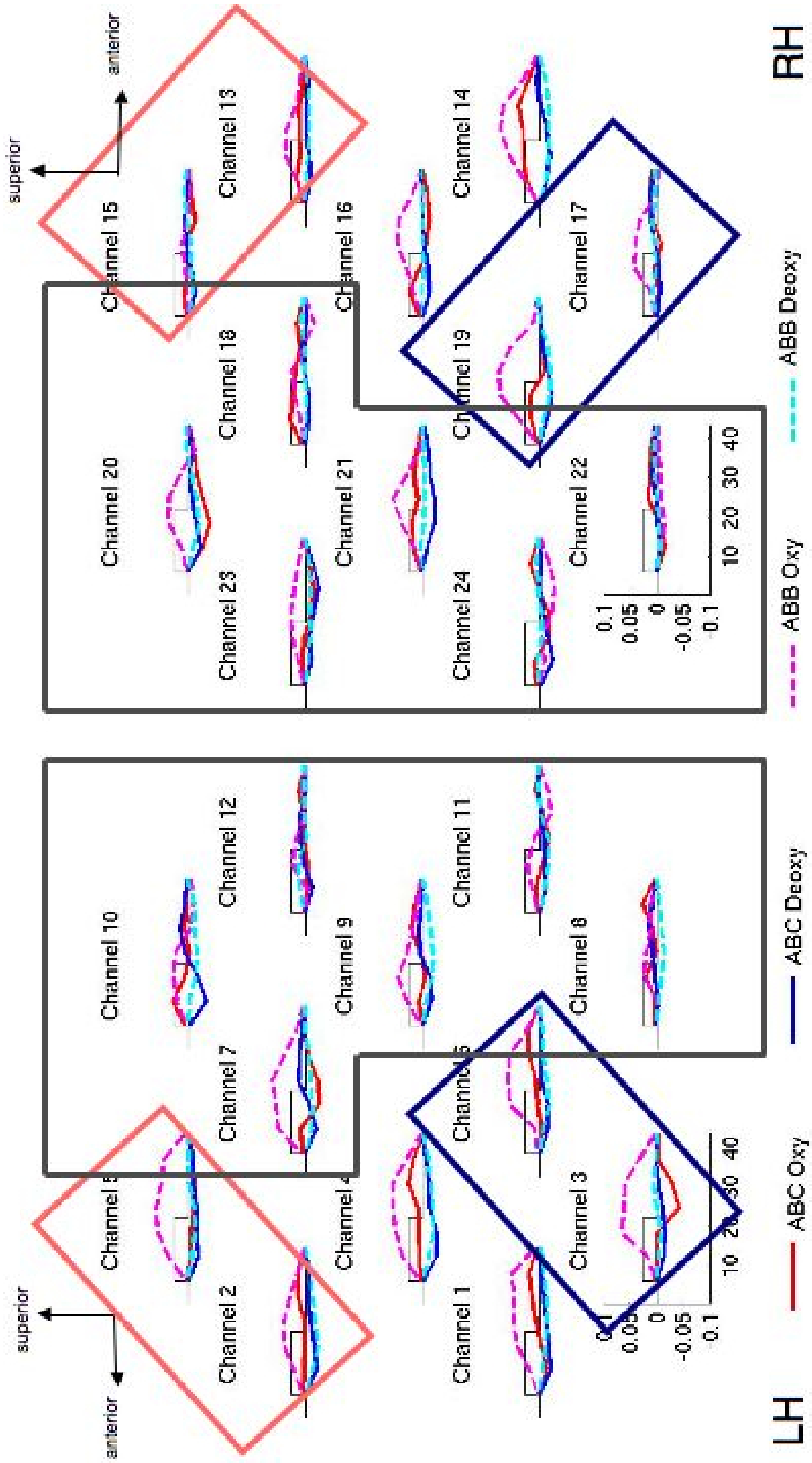
The resulting grand average of all neonates is shown in Figure 2.5. In a repeated measures ANOVA with factors Grammar (ABB/ABC), Hemisphere (left/right) and ROI (anterior/posterior) using oxyHb as the dependent measure, I obtained a main effect of Grammar ($F(1, 21) = 4.818, p = 0.040$) due to a greater overall activation for ABB than for ABC. The main effect of ROI ($F(1, 21) = 11.001, p = 0.003$) was also significant, the anterior regions being more activated than the posterior ones. In addition, there was a significant interaction between Grammar × Hemisphere ($F(1, 21) = 5.275, p = 0.033$), indicating larger activation for the ABB grammar in the LH. A similar ANOVA with factors Grammar (ABC/ABB) × Hemisphere (left/right) × ROI (anterior/posterior) was conducted for deoxyHb, and did not yield significant results, although the interaction Hemisphere × ROI showed a trend to significance ($F(1, 21) = 3.561, p = 0.073$).

It has to be noted that the responses to ABC appear to be relatively weak in certain channels. It has to be remembered, though, that the light intensity used in this experiment is only around 0.30–0.35 mW, which is much lower than the intensities used in other studies, e.g. it is half of the laser power used in the Peña et al. (2003) study.

Analysis of language-related areas

To get a better understanding of the mechanisms involved, I investigated auditory processing in the temporal areas and the processing of structure in the temporal areas. I ran an ANOVA with factors Grammar (ABB/ABC), Hemisphere (left/right) and Area (frontal/temporal), using oxyHb concentrations as the dependent measure (Figure 2.5). I obtained a significant main effect of Grammar ($F(1, 19) = 5.516, p = 0.030$), as before, due to a larger overall activation for the ABB grammar. No other main effect was significant. The interactions Grammar × Area ($F(1, 19) = 6.321, p = 0.021$) was significant due to larger activation in the temporal areas for the ABB than for the ABC grammar. There was also a significant Hemisphere × Area ($F(1, 19) = 6.603, p = 0.019$) interaction, due to larger activation in the left frontal than in the right frontal areas. The interaction Grammar × Hemisphere showed a trend to significance ($F(1, 19) = 3.094, p = 0.095$) due

Figure 2.5 *(following page)*: The grand average results obtained in Experiment 1. The positions of the channels correspond to the placement illustrated in Figure 2.4. LH appears on the left panel, RH appears on the right panel. Dark grey frames contain the channels of the posterior ROI. Unframed channels constitute in the anterior ROI. Blue rectangles enclose the temporal channels. Red rectangles enclose the frontal channels. The x-axes of the individual graphs represent time in seconds. Rectangle over the x-axes indicate time of stimulation within a block. The y-axes indicate concentration change in mmol·mm. The Hb concentration curves are color-coded in the following way: continuous red line: oxyHb concentration in the control ABC condition; continuous blue line: deoxyHb concentration in the control ABC condition; dashed pink line: oxyHb concentration in the test ABB condition; dashed turquoise line: deoxyHb concentration in the test ABB condition.


superior
anterior
Channel 15
Channel 13
Channel 16
Channel 14
Channel 17
Channel 18
Channel 19
Channel 20
Channel 21
Channel 22
Channel 23
Channel 24
Channel 10
Channel 12
Channel 9
Channel 11
Channel 8
Channel 7
Channel 6
Channel 5
Channel 4
Channel 3
Channel 2
Channel 1
0.1
0.05
0
-0.05
-0.1
10
20
30
40
ABB Oxy
ABB Deoxy
ABC Oxy
ABC Deoxy
RH
LH

to larger activation in the LH for the ABB than for the ABC grammar. A similar ANOVA for deoxyHb yielded no significant results.

Upon visual inspection of the data at the individual level, 11 of the 22 neonates showed frontal activation for ABB in the LH. Out of these, 8 also showed some frontal activation in the RH for ABB. No baby had frontal activation in the RH only. This frontal activation was closely tied to differential activation for ABB in the temporal areas. All the 11 babies who exhibited LH frontal response to ABB also showed differential activation to ABB in the left and right temporal AOIs, as well. No newborns showed activation in the temporal areas without activation in the frontal areas as well. This pattern of results, and the presence or absence of the (left) frontal activation, in particular, cannot simply be related to the state of alertness/sleep of the babies, since only about 2–3 of them were awake during the study. The others were in a state of sleep (although the actual sleep state was not monitored).

Analysis of time course

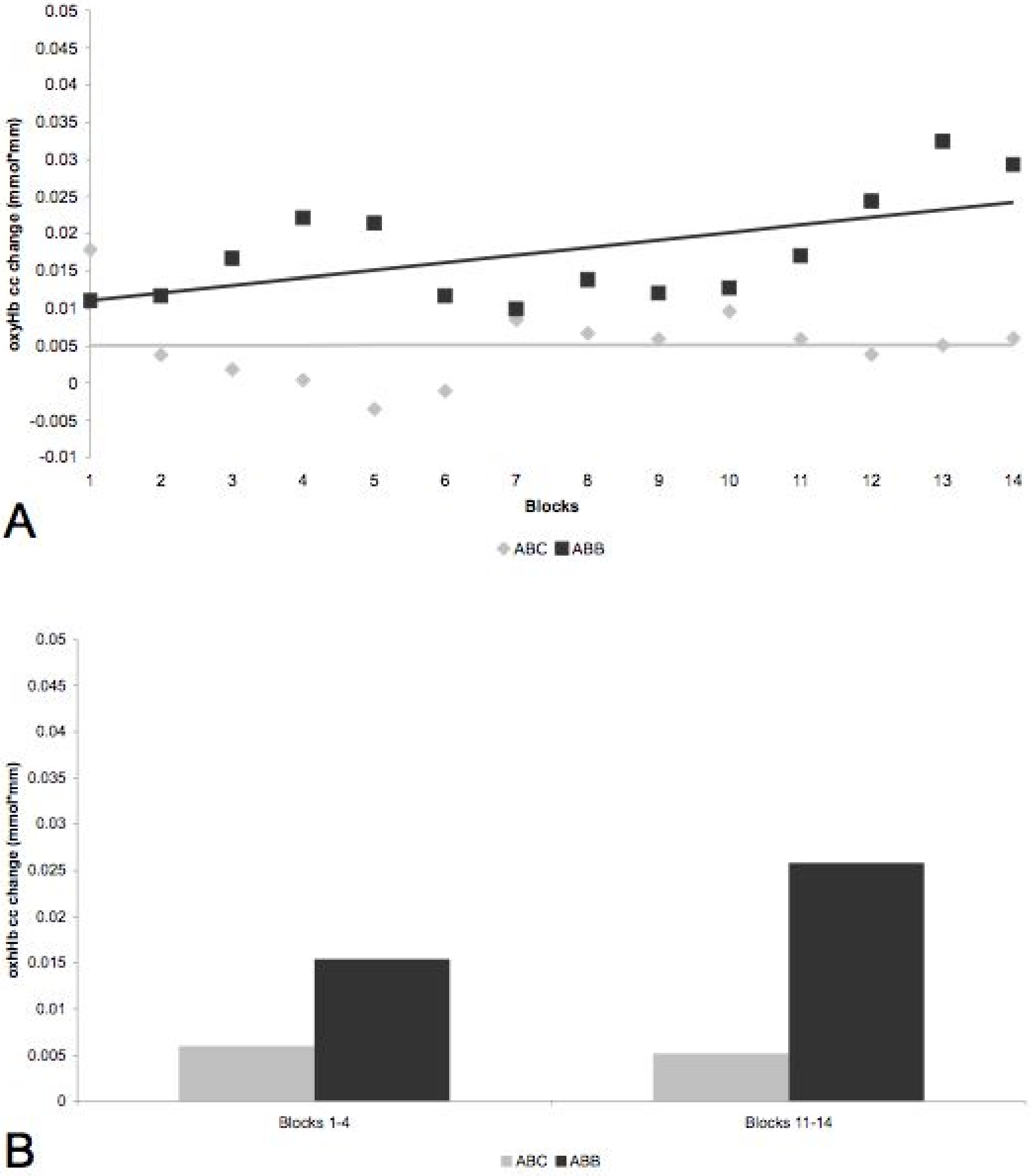
I also analyzed the temporal evolution of the responses during the course of the experiment in order to evaluate learning. Figure 2.6A illustrates the oxyHb concentration changes in the left anterior ROI over the 14 consecutive blocks of the experiment for the two grammars, as well as the linear regression line fitted on the learning curve. As indicated by the r^2 values (ABC: $r^2 = 0.00002$, ABB: $r^2 = 0.3427$), considerable learning only takes place for the ABB grammar. For statistical purposes, I compared the beginning and the end of the experiment, defined as the first and the last 4 blocks per grammar. Figure 2.6B illustrates the averages of the oxyHb concentrations in the two time periods for the two grammars. In an ANOVA with factors Grammar (ABC/ABB) × Time (beginning/end), I obtained a significant main effect of Grammar ($F(1, 21) = 7.174, p = 0.015$), as before, since the response to the ABB grammar was greater than the response to the ABC throughout the experiment. The main effect of Time was not significant. Importantly, there was a significant interaction between Grammar × Time ($F(1, 21) =$

$6.136, p = 0.023$). This was due to the fact that while the response to ABC tended not to change throughout the experiment, ABB elicited increasing activation over time, yielding an even greater difference between the two grammars towards the end of the experiment.

2.1.6 Discussion

First and foremost, these results indicate that the neonate brain is able to detect structure in the linguistic input. More specifically, it is able to detect and 'learn' a grammar containing adjacent repetitions. Importantly, this learning involves both a perceptual Gestalt-like configuration and a generalization mechanism. The repetition is detected as a 'perceptual primitive', as is evidenced by the significant difference between the two conditions already at the beginning of the experiment. Then, over the entire time course, further structural computations take place, extracting the underlying ABB generalization from the repetition sequences, which are all different from each other, but they all instantiate the same structural regularity. This 'learning' builds up over time, as the significantly increasing response to ABB shows. No such enhancement is observable for ABC. The enhanced response to ABB was mostly confined to the left anterior region, and in particular to left frontal areas. Since activation in these areas have been associated with grammar learning in previous experiments in adults (Friederici, 2002; Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006) and with working memory in older infants (Dehaene-Lambertz et al., 2006), the pattern of results obtained here can be interpreted as a brain signature for the processing of linguistic structure in neonates.

Figure 2.6 *(following page)*: The time course of the responses in Experiment 1. A: The linear regression lines of the oxyHb concentrations fitted on the data points provided by the 14 consecutive blocks for the two grammars. The light grey line represents ABC, the dark grey line represents ABB. B: The bars indicate the average oxyHb concentration in the left anterior region for the first and the last 4 blocks for the two grammars. The y-axis shows the average oxyHb concentration in mmol·mm. The light grey bars indicate ABC, the dark grey bars indicate ABB.


0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
-0.005
-0.01
oxyHb cc change (mmol*mm)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
Blocks
A
◆ ABC ■ ABB
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
oxyHb cc change (mmol*mm)
Blocks 1-4
Blocks 11-14
B
■ ABC ■ ABB

From a neurodevelopmental point of view, the observed LH superiority converges with previous results on language lateralization in (most) adults (Kimura, 1967) and infants (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002; Peña et al., 2003), indicating that the functional organization of the neonate brain is at least partially similar to that of adults.

2.2 Experiment 2: Non-adjacent repetitions (ABA)

The previous experiment has established that newborns' brains are endowed with repetition-detecting perceptual primitives, as well as a learning mechanism subserving structural generalizations. What is the nature of these mechanisms? What representations do they work on? Do they take any kind of repetition as input?

In natural language, dependencies hold not only adjacently, but also at a distance. Therefore, as a first step to explore the representations used by the two mechanisms identified above, in this experiment I test whether they can operate over non-adjacent repetitions (ABA), comparing it to the same ABC random control as before.

2.2.1 Material

The ABA grammar was derived from ABB by moving the first repeated syllable in each sentence into the initial position. All other parameters were kept identical. The ABC grammar was identical to the one used in Experiment 1.

2.2.2 Subjects

Another group of twenty-two healthy, full term neonates (12 females; mean age: 2.86 days, range: 2-5 days; Apgar ≥ 8) born to Italian-speaking families participated in Experiment 2. Four additional babies were tested, but not included in the analyses, because they failed to complete the experiment due to crying (3) or parental intervention (1). As before, parents gave informed consent prior to the experiment. The Ethics Committee of the Azienda Ospedaliero-Universitaria di Udine, where the experiments took place, granted permission.

2.2.3 Procedure

The procedure was identical to the one used in Experiment 1 with the exception that the power of the laser lights was 0.75 mW.

2.2.4 Data Analysis and Statistics

The analyses performed were identical to those conducted for Experiment 1 with the exception that, since laser power was increased, movement artifacts were now defined as concentration changes larger than 0.15 mmol·mm over a time span of 0.1 msec, i.e. 1 samples.

2.2.5 Results

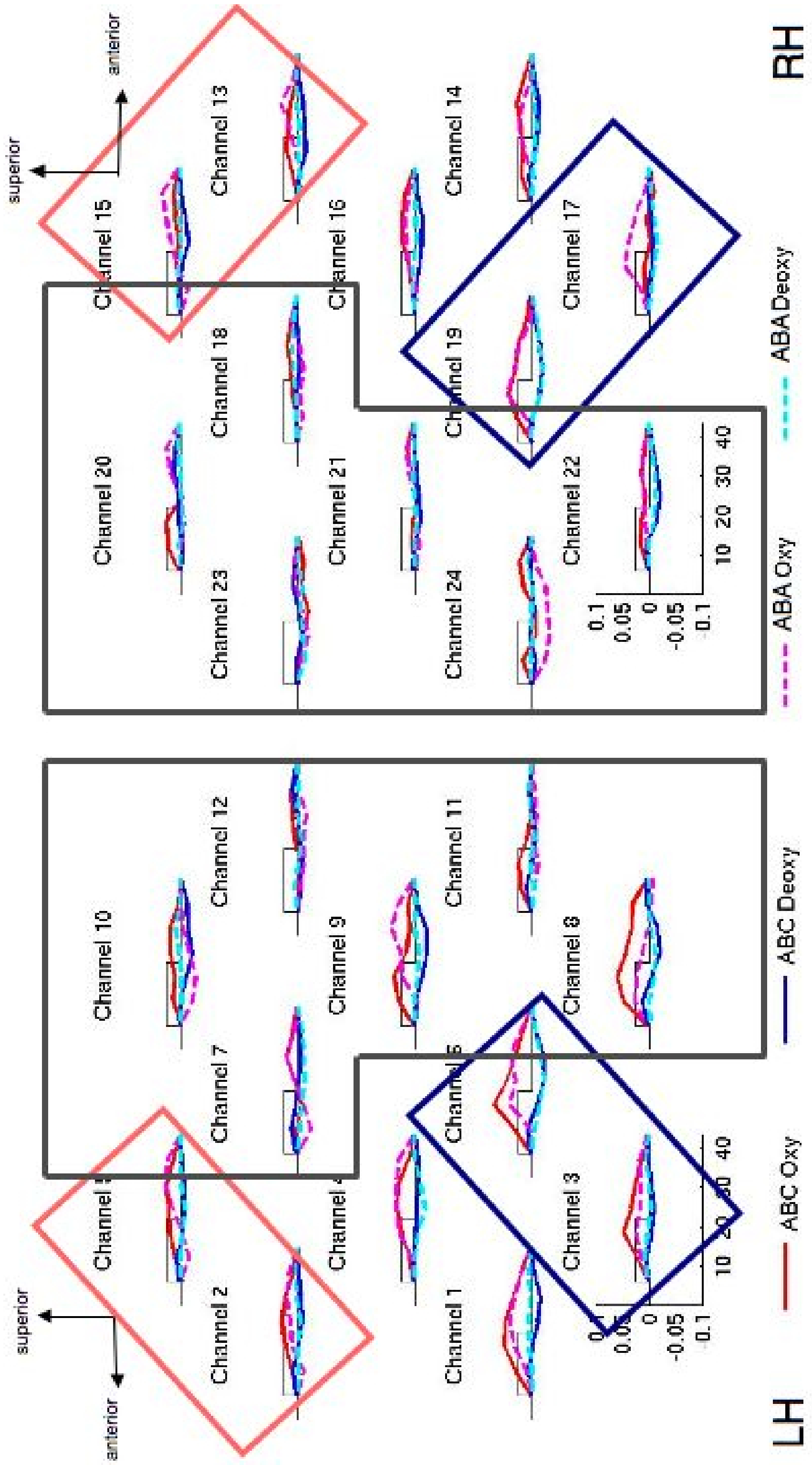
Overall analysis

The grand average results are shown in Figure 2.7. The ANOVA with factors Grammar (ABA/ABC), Hemisphere (LH/RH) and ROI (anterior/posterior), using oxyHb concentrations as the dependent measure, revealed no significant effect of Grammar or Hemisphere, but the effect of ROI was significant ($F(1, 21) = 11.470, p = 0.003$). No interactions were significant. A similar ANOVA using deoxyHb concentrations revealed no significant main effects or interactions.

Analysis of language-related areas

An ANOVA with factors Grammar (ABA/ABC), Hemisphere (LH/RH) and Area (frontal/temporal), using oxyHb concentrations as the dependent mea-

Figure 2.7 *(following page)*: The grand average results of Experiment 2. The same graphical conventions and color codes were used as for Figure 2.5.


RH
LH
superior
anterior
superior
anterior
Channel 1
Channel 2
Channel 3
Channel 4
Channel 5
Channel 6
Channel 7
Channel 8
Channel 9
Channel 10
Channel 11
Channel 12
Channel 13
Channel 14
Channel 15
Channel 16
Channel 17
Channel 18
Channel 19
Channel 20
Channel 21
Channel 22
Channel 23
Channel 24
0.1
0.05
0
-0.05
-0.1
10
20
30
40
ABC Oxy
ABC Deoxy
ABA Oxy
ABA Deoxy

sure, was conducted (Figure 2.7). I obtained a significant main effect of Area ($F(1,21) = 9.506, p = 0.006$), temporal areas being more activated than frontal ones. No other main effects or interactions were significant. A similar ANOVA for deoxyHb concentrations yielded no significant main effects or interactions.

Analysis of time course

Evaluating the time course of the responses in an ANOVA with factors Grammar (ABA/ABC) and Time (beginning/end), I found no main effects or interactions.

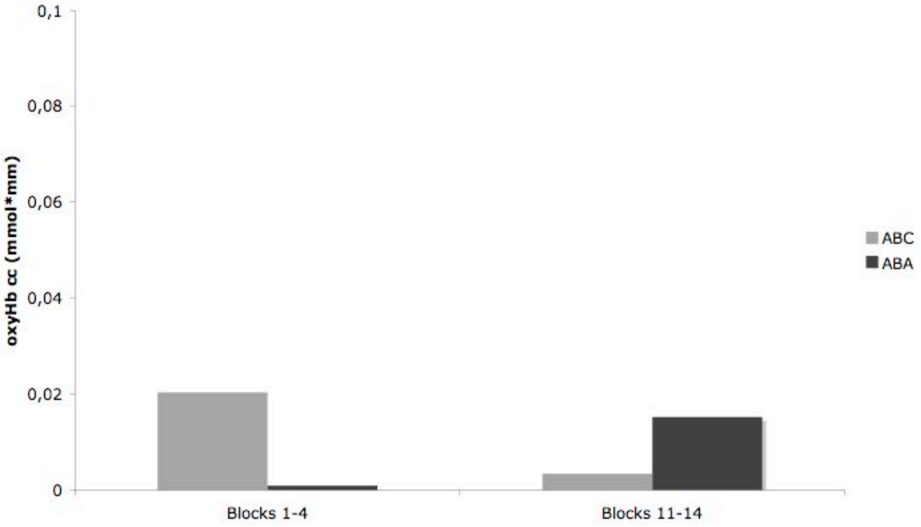


Figure 2.8: The time course of the responses in Experiment 2 (oxyHb). The same graphical conventions and color codes were used as for Figure 2.6B. In the absence of significant learning, regression lines for all 14 blocks are not shown.

2.2.6 Discussion

These results indicate that non-adjacent repetitions are not processed differently by the neonate brain than random sequences. Non-adjacent repetitions do not activate the repetition-detector, thus there is no output that could feed into the generalization mechanism. A grammar containing non-adjacent repetitions is processed as unstructured input. It elicited auditory processing, as witnessed by the significant activation of the temporal AOI, but no structural learning, as is evident from the lack of significant activation in the frontal areas and the absence of any change in the response over time.

Since both conditions yielded a simple auditory response in the current experiment, a comparison with the unstructured control condition of the previous experiment would be in order. It could provide a replication of the previous findings for the control grammars, which were identical in the two experiments. Additionally, it would further confirm the lack of structural processing in the non-adjacent grammar. Although such a statistical comparison would be necessary and informative, it cannot be performed due to an important procedural difference between the two experiments. As pointed out earlier, the light intensities used in Experiment 1 were around 0.30–0.35 mW, whereas those employed in Experiment 2 were two times higher, around 0.75 mW, due to an update of the NIRS machine. This difference precludes direct numerical comparisons between the experiments.

These findings lend further support to the proposal that adjacent repetitions are perceptually 'special', detected automatically by a dedicated perceptual component, rather than a symbolic computational mechanism. The latter should be able to compute the identity of any two symbols, irrespective of their distance, at least within the limitations of working memory.

But maybe the neonate brain's failure to detect and 'learn' non-adjacent repetitions is indeed a memory problem. It might be the case that it is not the intervening B syllable per se that interferes with the detection of the repetition, but that the newborn brain is simply unable to store the first A syllable for long enough to match it up with the later replica. The next experiment was designed to rule this possibility out.

2.3 Experiment 3: 'Non-adjacent adjacent' repetitions (A_A)

To tease the memory limitation account and the non-adjacency proposal apart, a new repetition grammar was created that poses no adjacency problem at the representational level, i.e. the repeated syllables are representationally adjacent, but they are temporally distant (A_A). If newborns' brains fail to detect this 'non-adjacent adjacent' repetition, then memory (or other performance factors) might be responsible. However, if these repetitions are detected, then adjacency is computed at the representational level.

2.3.1 Material

The A_A and A_C grammars was derived from ABA and ABC, respectively, by replacing the middle syllable with a pause of equal length (270 msec). Other parameters were left unchanged.

2.3.2 Subjects

21 newborns (6 females; mean age: 3.0 days, range: 2-4 days; Apgar ≥ 8) participated in this Experiment. Three more babies were tested, but were not included in the analyses, because they failed to complete the experiment due to crying. As before, parents gave informed consent prior to the experiment. The Ethics Committee of the Azienda Ospedaliero-Universitaria di Udine, where the experiments took place, granted permission.

2.3.3 Procedure

The procedure was identical to the one used in Experiment 2.

2.3.4 Data analysis and Statistics

The analyses were identical to the ones performed for Experiment 2.

2.3.5 Results

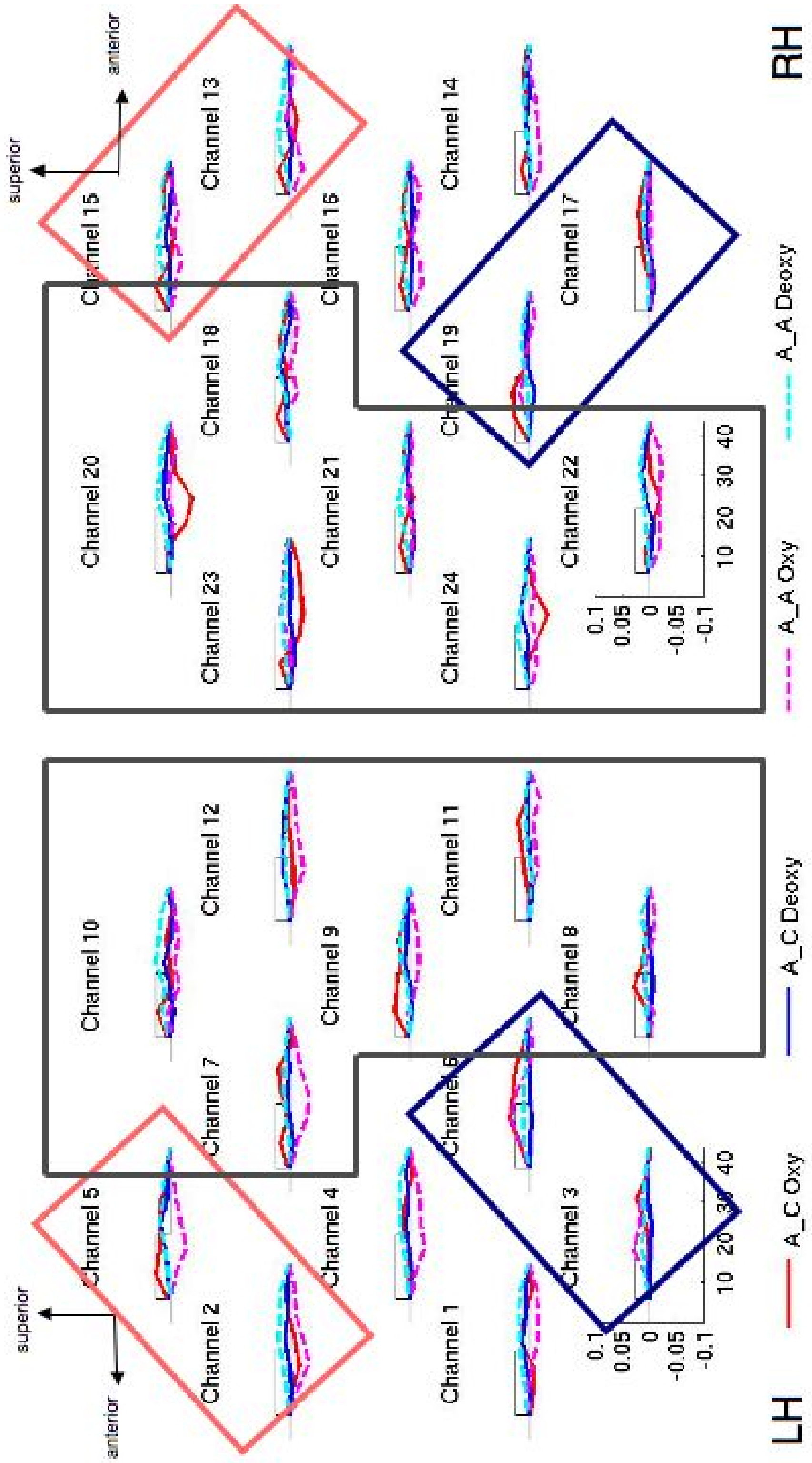
Overall analysis

The grand average results are shown in Figures 2.9 and 2.10 for a more convenient visualization. The ANOVA with factors Grammar (A_A/A_C), Hemisphere (LH/RH) and ROI (anterior/posterior), using oxyHb concentrations as the dependent measure, revealed no main effects or two-way interactions. The three-way interaction Grammar × Hemisphere × ROI was significant ($F(1, 20) = 6.530, p = 0.019$) due to a decrease in oxyHb concentration in the right posterior areas for the A_C grammar, while this grammar, unlike the A_A grammar, produced increased oxyHb concentrations in all other ROI. A similar ANOVA using deoxyHb concentrations revealed a significant main effect of Grammar ($F(1, 20) = 4.487, p = 0.047$). This reflects the fact that A_A gave rise to an increase in deoxyHb concentration, while A_C resulted in a decrease. No other main effect was significant. The two-way interaction Hemisphere × ROI showed a weak trend towards significance ($F(1, 20) = 3.073, p = 0.095$) due to a larger anterior than posterior activation in the LH, and the opposite pattern in the RH.

Analysis of language-related areas

In a more specific ANOVA with the factors Grammar (A_A/A_C), Hemisphere (LH/RH) and Area (frontal/temporal), using oxyHb concentrations as the dependent measure, I obtained a significant main effect of Area ($F(1, 20) = 9.127, p = 0.007$), because the temporal areas showed an increase, while the frontal areas exhibited no change or a decrease (Figure 2.11). The two-way interaction Grammar × Area was also significant ($F(1, 20) = 4.316, p = 0.050$). This reflects the fact that both grammars gave rise to increased activation in the temporal areas, whereas in the frontal areas, A_C induced practically no activity, while A_A produced a decrease in oxyHb concentra-

Figure 2.9 *(following page)*: The grand average results of Experiment 3. The same graphical conventions and color codes were used as for Figure 2.5.


LH
RH
superior
anterior
superior
anterior
Channel 1
Channel 2
Channel 3
Channel 4
Channel 5
Channel 6
Channel 7
Channel 8
Channel 9
Channel 10
Channel 11
Channel 12
Channel 13
Channel 14
Channel 15
Channel 16
Channel 17
Channel 18
Channel 19
Channel 20
Channel 21
Channel 22
Channel 23
Channel 24
0.1
0.05
0
-0.05
-0.1
10
20
30
40
0.1
0.05
0
-0.05
-0.1
10
20
30
40
A_C Oxy
A_C Deoxy
A_A Oxy
A_A Deoxy

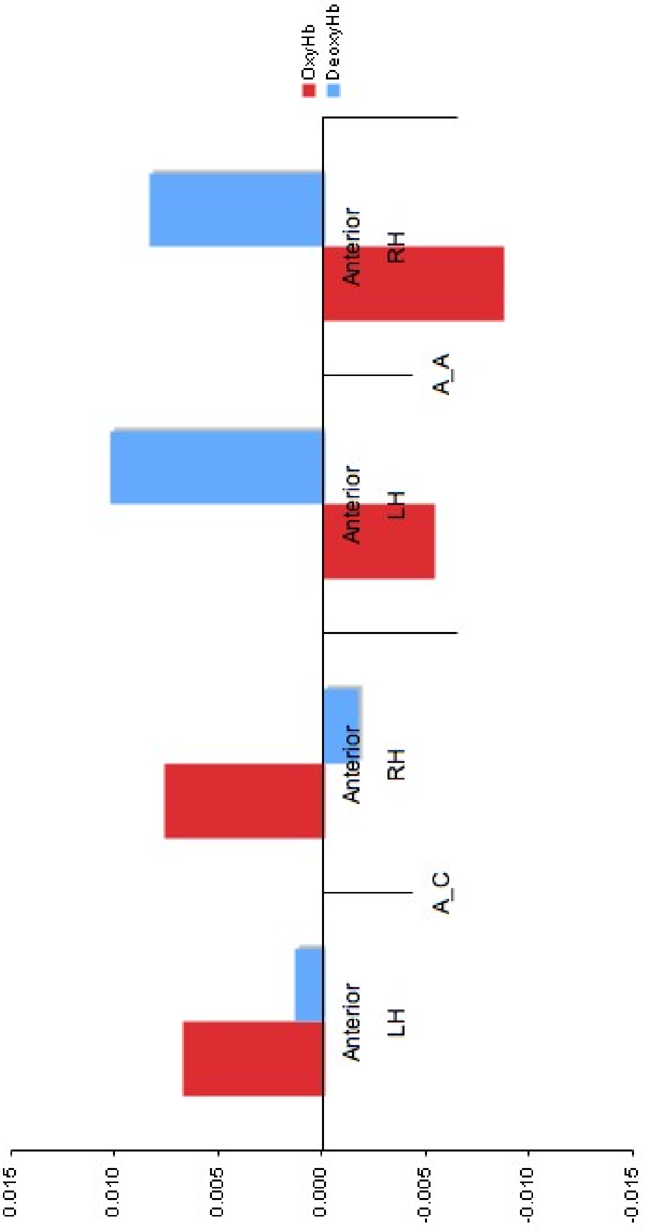


Figure 2.10: The grand average results of the anterior ROIs in Experiment 3, visualized as bar plots for convenience. These mean values were obtained by averaging over the 18 sec time windows of the stimulation for each ROI, as was done for the statistical analyses.

tion. In addition, there was a significant three-way interaction Grammar × Hemisphere × Area ($F(1,20) = 5.493, p = 0.030$) due to a larger decrease in the left frontal than in the right frontal areas, as well as a larger increase in the left temporal than in the right temporal areas for the A_A grammar, while the A_C grammar showed no hemispheric asymmetries. A similar ANOVA for deoxyHb concentrations yielded a significant main effect of Grammar ($F(1,20) = 4.884, p = 0.039$), reflecting the fact that A_A induced a large increase in deoxyHb concentrations, whereas A_C gave rise to decreased activation. The main effect of Area was also significant ($F(1,20) = 6.264, p = 0.021$), reflecting a larger increase in the frontal than in the temporal areas. No two-way interaction was significant, but the three-way interaction Grammar × Hemisphere × Area tended towards significance ($F(1,20) = 3.797, p = 0.065$) due to the fact that in the temporal areas, A_C produced decrease, A_A increase, particularly in the left, whereas in the frontal areas, A_C gave rise to practically no response, while A_A induced an increase, which was larger in the RH than in the LH.

Analysis of time course

Evaluating the time course of the oxyHb responses in the left anterior ROI in an ANOVA with factors Grammar (A_A/A_C) and Time (beginning/end), I found no main effects or interactions (Figure 2.12).

2.3.6 Discussion

The results indicate that the repetition-based A_A grammar was distinguished from the random A_C control. Therefore, the fact that the two identical syllables are separated in time does not interfere with the detection of repetitions. Consequently, the ABA grammar in the previous experiment failed to induce a distinctive response not because the A syllables were separated in time, but because a different B syllable intervened. This suggests that the simple memory limitation account can be excluded.

While the A_A repetition grammar was distinguished from the A_C control, the pattern of responses obtained were different than in the case of ABB

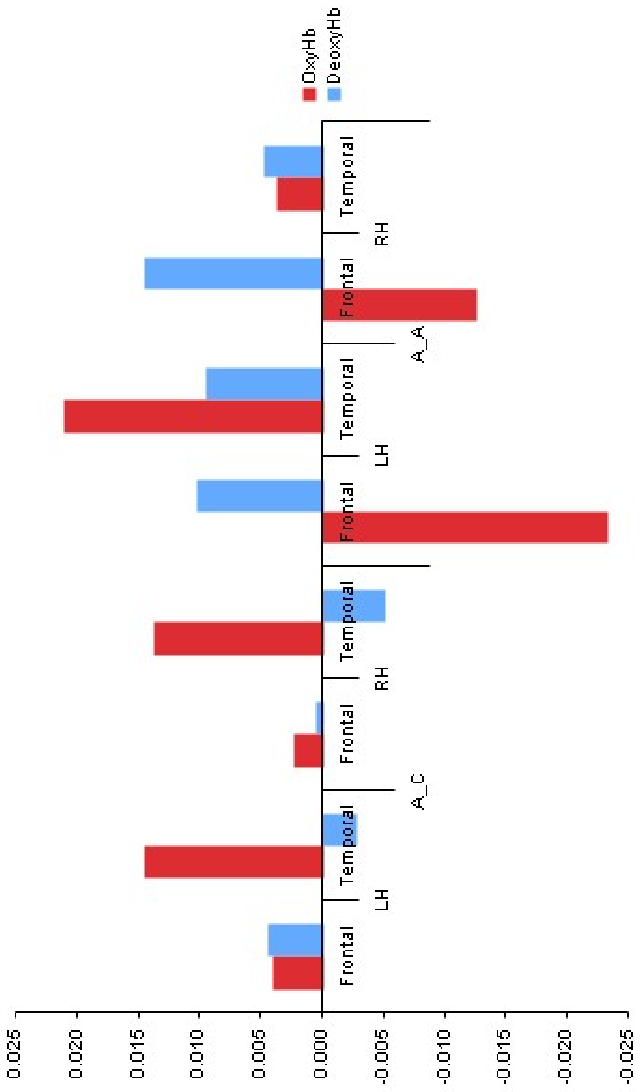


Figure 2.11: The grand average results of areas of interest in Experiment 3 visualized as bar plots for convenience.

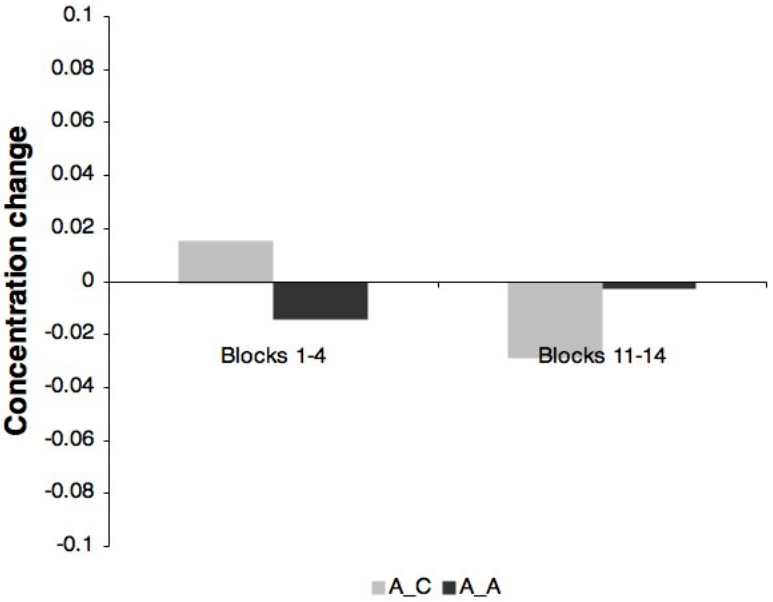


Figure 2.12: The time course of the responses in Experiment 3 (oxyHb). The same graphical conventions and color codes were used as for Figure 2.6B. In the absence of significant learning, regression lines for all 14 blocks are not shown.

vs. ABC in Experiment 1. Similarly to Experiment 1, both A_C and A_A induced an increase in oxyHb concentrations in the temporal, auditory areas. Unlike in Experiment 1, however, A_A gave rise to a large *decrease* in oxyHb concentrations and an *increase* in deoxyHb concentrations, i.e. an inverted response, in the frontal, structural areas. (The A_C control induced little frontal activation, just as in Experiment 1.)

The results indicate a clear dissociation between auditory processing and structural processing already in the neonate brain, suggesting early functional specialization in the processing of environing sounds (Dehaene-Lambertz et al., 2006; Friederici, 2002).

What is the reason for the inverted response in the structural areas? While the three experiments presented here cannot provide a definitive answer, certain explanations proposed in the literature to account for the inverted response can be excluded. First, it has sometimes been claimed (for a review, see Meek, 2002) that the direction of the response is subject to considerable inter-individual variation, and certain participants, especially infants, naturally show an inverted response. This is sometimes related to maturation (the excessive oxygen demand of synaptic proliferation and the immature vascular coupling) in the visual cortex of infants older than 5–8 weeks of age (Yamada et al., 1997; Morita et al., 2000). However, even these studies find classical, adult-like hemodynamic responses in newborns. Also, while vision matures considerably in infancy and early childhood (Mehler & Dupoux, 1994; Kovács, 2000), auditory perception is quite close to the adult state even at birth (Mehler & Dupoux, 1994). Furthermore, in the present experiment, such explanations can be excluded on empirical grounds, since within the same individual, we find a canonical response in one condition (A_C) and an inverted one in the other (A_A). Thus, variation obtains not between, and not within subjects.

Second, it might be argued that the state of alertness might play a role in the case of the OT signal. But, once again, this explanation is not adequate in this particular case, since within one individual, we find both response directions. Moreover, while I did not systematically monitor participants' state of alertness by physiological measures (heart rate, EEG etc.), observations during the experiment and the inspection of the videotapes indicate that the vast majority of babies (about 80-90%) were asleep throughout the experiment (although the specific sleep state cannot de determined).

Third, inverted responses are also found when a brain area is more active during 'rest' (participants quietly resting with the eyes closed) than during stimulation (Raichle et al., 2001). However, in the current study, there is no silence condition. Both the experimental and the control conditions involve auditory stimulation.

Fourth, increasing deoxyHb and decreasing oxyHb responses also appear when activation decreases in a pre-activated area of the brain. Wenzel et al. (2000) found inverted responses during ocular saccadic suppression. Obrig and colleagues (Hellmuth Obrig, unpublished data, personal communication) obtained classical hemodynamic responses during finger-tapping in the contralateral sensori-motor cortex due to contralateral inhibition. The same result was observed when a pulse was delivered to the relevant area of the motor cortex by transcranial magnetic stimulation (TMS) instead of active tapping. However, if the TMS pulse arrived when the area was already activated by finger-tapping, an inverted response was obtained. These results notwithstanding, the contextual modulation, e.g. sensitivity to pre-activation, of the hemodynamic response is not fully understood yet (see e.g. L. Gold & Lauritzen, 2002; Caesar, Gold, & Lauritzen, 2003; Caesar, Thomsen, & Lauritzen, 2003).

The selectivity of the inverted response in the current experiment suggests that some particular feature of the A_A stimuli is responsible. However, the exact effect of the temporal delay needs to be clarified before any definitive interpretation can be provided. To this effect, a fourth experiment is currently being run, comparing adjacent repetitions not embedded in a longer sequence, i.e. AA, with random controls, i.e. AC. If AA repetitions give rise to a response that is similar to what has been found for A_A, then the difference between ABB vs. A_A (and AA) is 'structural', possibly relating to the fact that in ABB the repetition is integrated into a lager unit. If AA patterns with ABB, then it is the immediate temporal adjacency of the repetitions that play a crucial role.

2.4 General Discussion: The role of perceptual primitives and generalizations in the initial state of the language faculty

A series of three experiments have been conducted to evaluate the presence of perceptual primitives and rule extraction in the initial state of the language faculty. Taken together, these findings, summarized in Table 2.2, show that upon its first encounters with language, the neonate brain is able to detect and learn structure from the input using both of these mechanisms.

	repetition	control	difference	response in frontal AOI	response in temporal AOI	learning
Exp 1	ABB	ABC	yes	classical	classical	yes
Exp 2	ABA	ABC	no	none	classical	no
Exp 3	A_A	A_C	yes	inverted	classical	no

Table 2.2: Summary of results obtained in Experiments 1–3. Column 'difference' indicates the presence or absence of a statistically significant difference between the two grammars. Columns 'frontal AOI' and 'temporal AOI' refer to the response in the repetition condition. Column 'learning' indicates a statistically significant difference over the time course of the experiment, i.e. the beginning and the end.

Since the neonate auditory system is already mature in most aspects, auditory perceptual primitives identified in adults (Endress et al., 2005) should also be observable at the youngest age. Indeed, repetitions have been shown to be detected when they appear adjacently (Experiments 1 and 3). Adjacency does not require temporal immediacy; the absence of intervening syllabic material is sufficient for the neonate brain to distinguish repetitions from random, but otherwise similar sequences. However, an intervening syllable renders the repetition undetectable.

The existence of such a repetition- or identity-detecting perceptual primitive raises a series of issues for the study of language acquisition. A set of questions concerns the notion of identity. From a philosophical point of view, it can be argued that identity cannot be inferred from experience. No two instances of a stimulus are exactly the same. Even if two objects constitute exact replicas in terms of their physical properties, e.g. two syllables that have the exact same waveform or spectrum, they will differ in the external circumstances in which they appear (e.g. Lewis, 1986 and related debates). Even if identical, syllable$_i$ will be produced at time t$_i$, while its replica at time t$_j$, or if produced simultaneously, then syllable$_i$ will be uttered by source$_i$, while its replica by source$_j$. It is not surprising, therefore, that an identity detector might exist as a primitive neural mechanism (for its existence in animals, see (Giurfa, Zhang, Jenett, Menzel, & Srinivasan, 2001). But even this logical consideration aside, the question arises what counts as an identical repetition for the neonate brain. The stimuli used in the above experiments contained acoustically (almost) identical reduplications.[2] But what if different instances of the same syllable are used? Or different speakers? Would the syllable /ta/ pronounced first by a female speaker, then by a male speaker count as a repetition? And do neonates normalize across variations in other acoustic properties, such as pitch or duration? Future experiments manipulating these properties of the stimuli might be able to provide us with a better understanding of the level of detail at which the neonate brain represents speech.

Why are non-adjacent repetitions not detected as a perceptual primitive? Two possible answers arise. First, non-adjacent repetitions might be perceptual primitives, but performance factors might limit neonates' ability to detect them. Experiment 3 has excluded one such limitation concerning memory span. However, it is possible that the limitation is of different nature, e.g. locality. If the identity-detector applies locally, the difference between ABA and ABC cannot be detected, because the relevant first and third syllables are never directly compared (ABA: A$\neq$B and B$\neq$A; ABC: A$\neq$B and B$\neq$C).

A second possibility is that non-adjacent repetitions are genuinely not

[2] Given that MBROLA (Dutoit, 1997) uses a diphone database for speech synthesis, the repeated syllables were not exact and perfectly identical replicas of each other, because each phoneme is co-articulated with its predecessor and successor. Thus the B syllables preceded by A were slightly different from B syllables preceded by B. However, these differences are minimal and undetectable by naive, untrained adults.

perceptual primitives. Adults are able to learn such patterns, because they can use mechanisms other than the perceptually based identity-detector, for instance symbolic operations (ABB: XYZ, where X=Z). Such symbolic operations might mature later in development, so they may not be available to neonates.

The three experiments systematically manipulated one of the perceptual primitives identified in adults (Endress et al., 2005, in press), namely repetitions. The other perceptual primitive was kept constant, i.e. repetitions always appeared at edge position (ABB, ABA, A_A). As a further step in understanding the function of perceptual primitives in language acquisition, it will be interesting to dissociate repetitions from edges in neonates and infants, as it has been done in adults (Endress et al., in press). Grammars could be constructed in which repetitions are removed further and further from the edges to investigate the role of positional codes (e.g. left edge/initial position, right edge/final position etc.).

The second mechanism identified by the above experiments is abstract rule extraction, which integrates the output of the identity detector into generalized structural representations. This mechanism is triggered when repetitions are initially detected as perceptual primitives (ABB, Experiment 1), but not when repetitions are not identified as such (ABA, Experiment 2). The latter sequences are processed as random ones, thus there is no higher level linguistic representation to learn.

Importantly for acquisition, rule extraction and generalization need to be further explored. Under what conditions does learning arise? What is the exact form of the abstract representation that emerges in neonates? In this regard, it is important to mention Carral et al.'s (2005) work, who also investigated simple auditory rule learning in newborns using ERPs. These authors found that the newborn brain could detect a deviant descending tone pair in a sequence of pairs of ascending tones. The tone pairs were all different from each other, so the differential response to the deviants had to be related to their deviant structure, not to their actual frequency.

The domain-selectivities of the perceptual primitive and the rule learning mechanism also need to be investigated. Are all auditory repetitions detected in a similar fashion? Will the ABB structure detected if it is implemented using pure tones or ambient noises? Recent experiments (Marcus, Fernandes, & Johnson, in press) suggest that tones, animal sounds and other ambient noises do not give rise to the discrimination and learning of ABB, AAB and ABA grammars in a (Marcus et al., 1999) paradigm. However, when syllables are used to teach the grammars during familiarization, just as in (Marcus et al., 1999), but in the test phase, grammars are implemented with tones, animal sounds or noises, knowledge is transferred and the grammars are successfully discriminated.

Is the repetition detector at work in other modalities? Would visual sequences with an ABB structure be distinguished from ABC sequences? Recent work from several laboratories (Saffran, Seibel, Pollak, & Shkolnik, in press; Gertner, Baillargeon, Marcus, Fisher, & Johnson, 2007) seems to suggest that ABB and ABA grammars can be learned in the visual modality, provided the stimuli are salient and natural enough. Interestingly, Gertner et al. (2007) also found an asymmetry between adjacent and non-adjacent repetitions. The authors used containers and occluders as visual stimuli to implement the ABB, AAB and ABA grammars. They found that ABB and AAB grammars were readily learned, while ABA was not.

From a neurodevelopmental point of view, the most important implication of these findings is that the functional organization of the neonate brain is already similar to that of the mature brain. A left hemisphere superiority is found for structured linguistic stimuli, just as in most adults (Kimura, 1967) and older infants (Dehaene-Lambertz et al., 2002). Moreover, speech, irrespectively of its structural properties, is processed by the temporal, perisylvian areas, implicated in auditory processing in adults, as well (Friederici, 2002). Linguistic structure, on the other hand, is computed in the frontal, in particular in the left frontal areas, which have repeatedly been found to engage in structure building and integration in adults (Friederici, 2002; Friederici et al., 2006) and older infants (Dehaene-Lambertz et al., 2006). While the localization of these areas remains somewhat imprecise in these experiments due to the low spacial resolution of NIRS and the lack of structural imaging to guide probe placement, our results accord well with previous findings and allows us to characterize the areas from a functional, rather than an anatomical, perspective.

Chapter 3

What is in the input? Computational analyses of infant-directed speech corpora in typologically different languages

The information contained in the input has been of central importance to most theories of language acquisition. According to learning-based theories, old (e.g. Skinner, 1957) and new (e.g. Tomasello, 2000), all that there is to language acquisition is the input and a general-purpose learning mechanism allowing the learner to match the input, especially its statistical properties, such as frequency distributions. Nativist theories took the opposite stance (Chomsky, 1959), drawing arguments precisely from the insufficiency of the input to guide acquisition (see Chapter 1).

While the nature of the input has played such a crucial theoretical role, relatively few studies have been conducted to critically assess the type of information it contains. Those few that have been carried out arrived at contrasting conclusions. Some (Yang, 2004; Gambell & Yang, 2004) argue that certain languages show morphosyntactic properties that preclude conditional probabilities à la Saffran, Aslin, and Newport (1996) from operating efficiently. For instance, languages in which words are predominantly monosyllabic, such as Chinese or (infant-directed) English, cannot be segmented at dips in transition probabilities (TPs)[1] between syllables, since most syllable boundaries are word boundaries at the same time, independently of the TP values. Others (Swingley, 2005) claim that some of these statistical computations, e.g. mutual information (MI) combined with frequency, can be useful even in English to build up a small initial vocabulary and to bootstrap later prosodic word learning strategies.

In this Chapter, I will examine infant-directed corpora in three typologically different languages, Japanese, Hungarian and Italian, to assess their information content, as measured by statistical cues commonly used in experimental psycholinguistics, namely conditional probabilities and frequency. These investigations cannot provide an exhaustive exploration of all the statistical properties of the input, but, at least, they probe some of their most relevant aspects.

I used infant-directed speech, which is different from adult-directed speech in a number of ways, including larger prosodic excursions, shorter utterances (Fernald et al., 1989; Jusczyk & Kemler Nelson, 1996), and a large number of identical repetitions (Sundberg, 1998). However, infant-directed speech is by no means a necessary prerequisite for language acquisition. In certain cultures, infants are addressed in the same language as adults or not even addressed very much at all (Gleitman, Newport, & Gleitman, 1984; Ochs & Schieffelin, 1984). Yet, even children in these cultures acquire their native language at the same pace as other children do. Infant-directed speech was chosen for the present purposes on the assumption that, although it is not necessary for acquisition, the properties in which it differs from adult grammar will play a role in acquisition.

In a first series of experiments, I evaluate whether statistical information, in particular forward transition probabilities (FWTPs), backward transition probabilities (BWTPs), FWTPS & BWTPs combined, and mutual information (MI), indeed provide useful cues for segmentation in Hungarian and

[1] For abbreviations, see p. 1.

Italian. Since these languages are morphologically different languages, I also ask the question whether segmentation as cued by conditional probabilities, is sensitive to morphological type, i.e. agglutination vs. inflection.

In two subsequent experiments, I investigate whether frequency, a simpler statistical measure, suffices to cue word categories, such as functors and content words, and through them, a basic typological property, i.e. word order, in Japanese and Italian.

Before I report the experiments, I review some of the existing work on statistical information contained in the input. Then, after a short detour to introduce the relevant linguistic properties of my target languages, I reformulate the research questions with greater linguistic specificity.

3.1 The status of the input in language acquisition

3.1.1 Language statistics: from information theory and structural linguistics to experimental psychology

At the very foundations of information theory lies the observation (Shannon, 1948) that the statistical structure of natural language, conceived of as a discrete symbolic system, is such that its units are neither equiprobable nor independent of each other. Therefore, given a string of units, it is possible to predict the upcoming unit with a probability that is different from chance. One of the measures that Shannon introduced to characterize this statistical structure is the probability of a unit x to appear after the sequence y of n number of units. This is known as the transitional probability (TP) from y to x. Very often n is chosen to be 1, so the most commonly used TP is the one between neighboring units.

These information theoretic notions immediately made their way into psychology. In a seminal paper, Miller (1956) established the principles of information processing and storage in the human mind. He proposed that a key notion in how the mind organizes information is 'chunking'—information is broken down into units the size of which ("the magical number seven, plus or minus two") is determined by the storage capacity of 'immediate memory'.

Faced with the problem of providing a valid morphological analysis of previously undescribed native American tongues *without* relying on native speaker intuition, structuralist linguists also developed statistical and distributional methods to identify the phonemes and morphemes of these unknown languages. Specifically, Harris (1955) proposed a way to establish morpheme boundaries in unsegmented utterances based on the intuitive idea that distributional coherence is stronger between phonemes that fall inside the same morpheme than between those that span morpheme boundaries. He suggested a method that counts, for all phoneme sequences of length *n* starting from the beginning of the utterance, the number of phonemes that appear in the *n+1*th slot in other utterances that start with the same phoneme sequence. For example, in the utterance *He's clever* /hiyzklevər/[2], the initial /h/ can have 9 successors in other English sentences, the sequence /hi/ can have 14, /hiy/ 29, /hiyz/ 29, /hiyzk/ 11, /hiyzkl/ 7, /hiyzkle/ 1, /hiyzklev/ 1, and /hiyzklevə/ 1. Morpheme boundaries are posited where the successor counts are the highest, i.e. their is a larger variability of potential successors—in this example, after /hiy/ and /hiyz/, where basically all English phonemes can appear as successors. This readily corresponds to the expected segmentation of the utterance into the morphemes *he, 's, clever.*

A necessary technical limitation of Harris's work is that at the time it could not be tested on large corpora, thus we are left with a few examples, such as the one above, as empirical evidence. Nevertheless, Harris discussed a number of issues to be taken into consideration when the method is to be applied to real data. He noted that successor counts performed poorly in two cases: (i) with ambiguous strings (like /ðeyl/, which can be segmented both as *they'll* and *they l[eft]*) and (ii) bound morphemes (prefixes, suffixes etc.), i.e. inside morphologically complex words. To overcome the former difficulty, Harris suggested the use of a more detailed phonological transcription, since allophony, co-articulation, stress and prosody very often disambiguate phonemically similar strings. In other words, the chosen level of representa-

[2] I am using Harris's (1955) original phonemic transcription.

tional detail has a non-negligible effect on the success of segmentation. Harris also claimed that both problems might be alleviated if the algorithm is applied not only forward, but also backward, and breakpoints are posited where both successor *and* predecessor counts are high. This is especially efficient in finding bound morphemes, and thus is the preferred segmentation strategy in languages with complex morphology, such as Telugu, which Harris briefly analyzes. Additionally, he suggested that using a larger successor/predecessor window, e.g. *n+2*, and taking into account not only successor/predecessor counts, but also the relative frequency of successors/predecessors can improve segmentation results.

Concerned with providing a quantitative, thus operational definition for the notion of word, Gammon (1969) adapted Harris's (1955) method to detect word boundaries in morpheme sequences as input representations. He took into account both successor and predecessor counts for each member in a morpheme pair, thus obtaining four measures for each boundary. The more of them converged towards indicating a word boundary, the higher the probability of the correctness of the resulting segmentation. Also, the more likely it was that the established word boundary coincided with a syntactic phrase boundary. Thus, the method could also be used for syntactic bracketing. As empirical evidence, Gammon reported segmentation results obtained with three small corpora (a few thousand morpheme tokens), all of them excerpts from literary texts. He concluded that segmentation was not so much affected by absolute sample size as by morphemic diversity (one of his texts was highly repetitive, the other two were more variable). In other words, segmentation is impaired by sparse data. In addition, Gammon noted the existence of a left-right asymmetry. If considered alone, successor counts had proven better predictors than predecessor counts (but, of course, the best segmentation was achieved when the two were combined).

As the problem of segmentation made its way from information theory and linguistics into psychology and machine learning theory, several computational studies (e.g., Kay, 1973; Olivier, 1968; Wolff, 1975, 1977) had been conducted with the aim of developing efficient algorithms. These studies mostly worked with artificially generated texts or very small natural language samples, and were little concerned by the statistical properties of real, natural input. Nevertheless, some of their conclusions are relevant for the present purposes. Inspired by early experimental results about adults' segmentation abilities (J. Hayes & Clark, 1970), Wolff (1975, 1977) described three segmentation algorithms and tested them on several artificial and natural language samples. One of the algorithms used forward TPs, the other two calculated co-occurrence frequencies/joint probabilities. The crucial difference between the two measures is that TPs are directional, thus asymmetric, while joint probabilities are symmetric. All three algorithms posited a boundary where the measures achieved a preset threshold: a TP lower than 0.25 or a co-occurrence frequency larger than 10. Wolff (1975) concluded that the TP algorithm performed less well than the joint probability one, unless absolute frequency thresholds were introduced as additional criteria. He also observed that joint probability algorithms achieved better than chance results even on natural language samples, but required a relatively large number of iterations. Wolff (1977) also briefly remarked that joint probability algorithms were sensitive to morphological structure inside complex words (e.g. they were able to segment out the plural or 3rd person singular -*s*). Thus, he proposed that on the basis of some measure of the strength of association between neighboring units, a graded, rather than a categorical segmentation could be achieved, reflecting morphological structure—e.g., word boundaries might show weaker association than do word internal morpheme boundaries, which, in turn, might be weaker than stem internal syllable pairs. Contrary to Gammon's (1969) findings, Wolff observed that his algorithm could not identify larger syntactic phrase boundaries efficiently.

When behavioral experiments on segmentation gained new momentum (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996), computational models also multiplied (e.g. Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997; Christiansen, Allen, & Seidenberg, 1998; Batchelder, 2002; Swingley, 2005; Gambell & Yang, 2004).

Brent and Cartwright (1996) proposed a model that segments the input using a Minimum Description Length (MDL) algorithm. This approach treats segmentation as an optimization problem. Accordingly, the algorithm creates all the possible segmentations of the input, and finds the one that minimizes the length of the full representation. A full representation consists of (i) a lexicon, i.e. an indexed list of all the word types that can be found in the segmentation, and (ii) a derivation, in which all the word tokens of the input are replaced by the index that the word type they belong to gets in the lexicon. The MDL algorithm returns the segmentation whose full representation (lexicon+derivation) is the shortest, i.e. consists of the smallest number of characters. Brent and Cartwright ran their algorithm on a small corpus of infant-directed speech (roughly 5000 words) from the CHILDES database, represented as a sequence of phonemes. When the possible segmentations were not phonotactically constrained, the algorithm achieved 41.3% accuracy and 47.3% completeness, which was significantly better than the performance (13.4% accuracy and 13.4% completeness) of the random segmentation algorithm that they used as a baseline. However, the algorithm was so computation-intensive that a larger corpus could not have been tested, as the authors themselves noted. Therefore, it is hard to see how such an optimization algorithm translates into cognitive mechanisms that are able to process the natural input, which is several orders of magnitude larger. Furthermore, the authors reported practically no qualitative results, so it is difficult to assess the segmentations from a linguistic point of view.

Batchelder (2002) introduced a different algorithm, called BootLex, which treated segmentation as a problem of finding the most likely parse of the input, given an up-datable lexicon of words and their frequencies. The routine parses the input on the basis of the current state of the lexicon, computes the likelihood of the possible parses from the word frequencies, then updates the lexicon with new word candidates obtained from the most likely parse. Word candidates are pairs of adjacent words that do not appear in the lexicon yet. Then the algorithm is re-initiated on the next utterance. To stop excessively long words from being entered into the lexicon, maximum word length is constrained. Batchelder tested the BootLex algorithm on different versions of six corpora, two English, two Spanish and two Japanese. One in each language was obtained from infant-directed utterances of the CHILDES database, the other contained written material from children's books or adult-directed language. Different encodings of each, e.g. phonetic, orthographic, hiragana etc., were used. Utterance boundaries were retained in all of the samples. The best performance, around 70% accuracy and completeness, was achieved with a large, phonologically transcribed English corpus of infant-directed speech. The lowest, around 30% accuracy and completeness, was obtained for a large Spanish orthographically coded corpus of written language. According to the author, the reason for the difference is that the English corpus contained shorter utterances, and consequently a higher number of utterance boundaries, shorter words and more repetitions than the other samples. In general, the Spanish and Japanese results always remained inferior to the corresponding English ones. It seems that the BootLex algorithm is hindered by the morphological complexity of the language.

The work of Swingley (2005) more directly addressed the question whether simple statistical computations are useful to the learner. He calculated absolute frequencies for each mono-, bi- and trisyllable of his English and Dutch corpora (42 000 and 26 000 words, respectively), plus mutual information, a symmetric condition probability measure (see Section 3.2.2 for a definition) for all bisyllables. A monosyllabic word was added to the lexicon if its frequency exceeded a threshold. A bisyllabic word was posited if both its frequency and the mutual information between the two syllables reached a threshold. A trisyllabic lexical entry was created if its frequency and the mutual information between the two constituting syllable pairs exceeded a threshold. The highest performance, about 80% accuracy, was achieved when the threshold was set to the 70th-80th percentile of the frequency and the mutual information ranks. However, even at this range, only about 300-400 words could be learned by the algorithm out of the 1800 English and 1050 Dutch word types. Thus, completeness, although the author didn't report exact figures, appears to be rather low. The final conclusion of the paper is that statistical learning might be useful to build up a small initial vocabulary, over which certain phonological generalizations could be made leading to further word learning strategies. It is interesting to note that Swingley used both frequency *and* mutual information. Although he did run versions of the algorithm with only one of the cues at a time to tease apart their relative contributions, he didn't report comparative statistics between the performances of the 'frequency only', the 'MI only' and the original, combined algorithm; nor did he provide any qualitative comparisons. The reported raw data seem to suggest, though, that frequency contributed much more to the original results than did mutual information.

Yang and his colleagues (Gambell & Yang, 2004; Yang, 2004) took yet another approach. Their goal was to establish a psychologically plausible segmentation mechanism. Therefore, given empirical data (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996) that infants can compute TPs over syllables, Gambell and Yang (2004) implemented an algorithm that calculated TPs on-line, i.e. counts were updated after each incoming syllable. Boundaries were posited at local TP minima. As input, they used infant-directed speech from the CHILDES database, phonologically transcribed and syllabified (226 178 words, 263 660 syllables). They obtained 41.6% accuracy and 23.3% completeness, a relatively poor result. The authors attributed this to the monosyllabicity of the input (a monosyllable follows a monosyllable 85% of the time), as the local minimum algorithm crucially assumes that words are multisyllabic. Moreover, as the authors reported, monosyllabicity remained a problem irrespectively of the size of the corpus; after the first 100 000 syllables were processed, TP values stabilized and changed very little on further input. Therefore, the authors concluded that "statistical learning is ineffective", and proposed alternative segmentation strategies based on phonological and algebraic regularities instead.

3.1.2 The main typological properties of Japanese, Hungarian and Italian

Since it is controversial how reliable segmentation cues statistics can offer, and whether the cues are equally informative in all languages, I have chosen two non-Indo-European languages, Japanese and Hungarian, as well as an Indo-European language that has not been studied from a statistical perspective yet, Italian, as a testing ground. For a better understanding of the specific segmentation and bootstrapping problems they pose, I present their main morphosyntactic characteristics below.

Japanese

Japanese is a Japonic language. It is characterized by a Complement-Head order, e.g. OV order (1), and complementizers that follow their subordinate clause (2)[3]. As is also clear from the examples, Japanese has agglutinating morphology, with a complex case system and a verbal inflection system that marks tense, modality etc., although it lacks person and number agreement. It has simple syllable structure, allowing only V, CV, and CVC* syllables, where C* can only be a nasal or the first half of an obstruent geminate consonant (see also Chapter 1 for more discussion and examples).

(1) *Taroo ga* ***tegami o*** *kaita.*
Taroo.NOM letter.ACC wrote
'Taroo wrote a letter.'

(2) *Mary ga* [*John ga hon o yon da* ***to***] *omottei ru*
Mary.NOM John.NOM book.ACC read.PST that think.PRES
'Mary thinks that John read a book.'

Hungarian

Hungarian is a Finno-Ugric language. Diachronically, it had a consistent Complement-Head order, and has preserved this property in many of its phrase types, e.g. OV order in non-perfective VPs ((3a)), postpositions ((4)), prenominal relative clauses etc. The VP, however, synchronically also exhibits a VO pattern (3b)[4], namely when the verb is perfective. Hungarian is

[3] Abbreviations:
NOM: Nominative case; ACC: Accusative case; PRES: present tense; PST: past tense

[4] Abbreviations:
PRT: perfective preverbal particle

agglutinating, with heavy nominal (5)[5] and verbal (6)[6] morphology. Also as a result of historical development, Hungarian allows fairly complex syllable structures: (C)(C)(C)V(C)(C), e.g. *sport* 'sport', *struktúra* 'structure'.

(3) a. *Almát eszik.*
apple.ACC eat.3SG
'He is eating an apple.'

b. *Megeszi az almát.*
PRT.eat.3SG the apple.ACC
'He is eating up (all) the apple.'

(4) *az asztal alatt*
the table under
'under the table'

(5) a. *ház*
house.NOM
'house'

b. *házat*
house.ACC
'house'

c. *házhoz*
house.ALL
'to (a/the) house'

d. *házba*
house.ILL
'into (a/the) house'

e. *házban*
house.INE
'in (a/the) house'

f. *házból*
house.ELA
'from (a/the) house'

g. *házakból*
house.PL.ELA

[5] Abbreviations:
ALL: Allative case; ILL: Illative case; INE: Inessive case; ELA: Elative case; PL: plural; 1SGPOSS: 1st person singular possessive

[6] Abbreviations:
3SG: 3rd person singular; 1PL: 1st person plural; MOD: modal

‘from (a/the) houses’

h. *házaimból*
house.PL.1SGPOSS.ELA
‘from my houses’

(6) a. *lát*
see.3SG
‘he sees’

b. *látunk*
see.1PL
‘we see’

c. *látott*
see.PAST.3SG
‘he saw’

d. *láttunk*
see.PAST.1PL
‘we saw’

e. *láthattunk*
see.MOD.PAST.3SG
‘we could see’

Italian

Italian is an Indo-European language. Its word order is Head-Complement, e.g. VO (7), prepositions (8). It has inflecting nominal (9) and verbal (10) morphology. In its syllable structures, complex onsets are allowed, but (complex) codas are dispreferred: (C)(C)(C)V(C), e.g. *stretto* ‘tight, narrow’.

(7) *Gianni prende un caffè.*
Gianni takes a coffee
‘Gianni is having a coffee.’

(8) *sotto il tavolo*
under the table
‘under the table’

(9) a. *ragazzo*
boy.SG
‘boy’

b. *ragazzi*
boy.PL
'boys'

c. *ragazza*
girl.SG
'girl'

d. *ragazze*
girl.PL
'girls'

(10) a. *arrivo*
arrive.1SG
'I arrive'

b. *arriva*
arrive.3SG
'he arrives'

c. *arrivai*
arrive.PAST.1SG
'I (have) arrived'

d. *arrivò*
arrive.PAST.3SG
'he (has) arrived'

The three languages represent a wide range of typological options, allowing to test the informativeness of statistical cues more universally, extending the scope of investigations to languages with heavy morphologies.

3.1.3 Developmental questions: what can statistical information cue?

Given such languages, we can test some highly specific questions about the role of statistical information in language acquisition.

First, are the boundaries of (morphologically complex) words reliably signaled by some statistical measure(s)? This is what most previous studies on segmentation, both computational and behavioral, addressed. In Experiment 4, I extend the question to Hungarian and Italian.

Second, are the boundaries of morphemes inside complex words reliably signaled by some statistical measure(s)? Are these the same as the ones that signal word boundaries? This a morphosyntactic extension of the first question, already raised by some authors (Harris, 1955; Gammon, 1969; Wolff, 1977; Antal, 1977). Since they mostly worked with small English corpora, they could only provide partial evidence. Therefore, Experiments 4 and 5 address this question using larger corpora in the two morphologically rich target languages.

Note that the first two questions are logically independent of each other. It might turn out to be the case that the statistical properties of the input indicate only morphologically complex word boundaries—providing an empirical basis for the intuition that *asztalra* (table-onto 'onto the table') is a word in Hungarian, but the suffix *-ra* in isolation is not. Conversely, statistics might only signal morpheme boundaries, independently of whether the morphemes are free (*asztal*) or bound (*-ra*), thus providing input for word learning and the lexicon, and leaving it to the morphological and syntactic components to combine morphemes into what are intuitively known as words. Ultimately, these issues might shed some light on the complex interactions between the lexicon and the morphosyntactic component.

Third, can some statistical measure(s) help the child boot-strap the basic morphosyntactic properties of the target language? Does a given statistical measure show different distributions in typologically different languages, thus cuing morphosyntactic type? These questions formulate a novel way of looking at statistical information. It is not a new idea that distributional analysis plays a role in the acquisition of syntax, e.g. by establishing syntactic categories (Harris, 1951; Mintz, Newport, & Bever, 2002; Mintz, 2003). However, I am raising the possibility of a different kind of interaction between statistical distributions and the acquisition of syntax. The question I am asking is whether statistics can act as a surface cue to the acquisition of some of the most general typological properties of the target language.

There are at least two possibilities worth considering in this regard. First, if the answers to the first two questions raised above turn out to be positive, statistics will provide a cue to distinguish between functionally very similar phrases, e.g. case-marked nouns (*asztalra*) and postpositional phrases (*asz-*

tal alatt, table under, 'under the table'). Distinguishing such morphosyntactic constructs provides evidence for separating functionally similar free and bound morphemes (postpositions and suffixes, respectively). This, in turn, indicates the agglutinating nature of the target language, and provides an important cue about Head-Complement order.

But even if the answers to the two initial questions turn out to be negative, i.e. statistical distributions do not provide reliable enough information about segmentation, they might still offer a global cue to the morphosyntactic type of the language. To put it differently, statistical measures may not be precise enough to signal morpheme or word boundaries, but their distributions in morphologically different languages might still be sufficiently different to indicate morphological type. In this sense, statistical distributions could work as a bootstrapping cue, somewhat similarly to %V, the amount of vocalic space in the speech signal (Ramus & Mehler, 1999; Ramus, 2002), which correlates with morphological type and basic word order (Mehler et al., 2004). Experiments 4 and 6 explore these possibilities.

3.2 Experiment 4: the role of statistical cues in segmenting Hungarian and Italian infant-directed speech

This experiment consists of a series of studies that systematically investigate the efficiency of a number of statistical measures in segmenting words and morphemes in Hungarian and Italian infant-directed corpora and in signaling the morphological properties of these languages.

As is evident from the review of the literature, many different statistical measures and segmentation algorithms exist. Of these, I used the most common ones: forward transition probability (FWTP), backward transition probability (BWTP), the combination of FWTPs+BWTPs and mutual information (MI) as statistical measures, and absolute as well as relative thresholding as segmentation algorithms. All statistical measures were combined with all segmentation algorithms.

The four conditional probability measures were chosen in order to explore which one is able to capture the regularities of agglutinating morphology the best. As previous results suggest (Harris, 1951; Gammon, 1969), the joint use of both FWTPs and BWTPs might yield good segmentation at the morpheme level. Specifically, if TPs are useful for segmentation, the following pattern might be expected:

		FWTPs	
		low	high
BWTP	low	word boundary	suffix
	high	prefix[7]	root internal

Thus, while the use of FWTPs or BWTPS alone yield one type of boundary, their joint use (henceforth, FWTPs+BWTPs) allows to posit two types of boundaries. Transitions of the lowest statistical coherence (low FW and BW TPs) can be interpreted as word boundaries (WB), transitions of intermediate coherence (low FW and high BW TPs or high FW and low BW TPs) as word internal morpheme boundaries (WIMB), while transitions of high coherence constitute no break points at all.

Another way of combining forward and backward predictability is to use a conditional probability measure that is symmetric and incorporates probabilities from both directions. I have chosen MI among the symmetric probability measures. This way, the effects of computing FWTPs and BWTPs separately and then combining them can be compared to directly calculating a symmetric measure.

More complex information theoretic measures, e.g. mean description length (MDL), are sometimes also employed in the literature, mostly in studies that treat segmentation as an optimization problem (e.g. Brent & Cartwright, 1996). I have not used any of these in my analyses, since there is no empirical evidence that humans are able to compute them to solve segmentation tasks.

3.2.1 Corpora

I used a Hungarian (MacWhinney, 1974, 1975; Réger, 2004) and an Italian (Antelmi, n.d.; Antinucci & Parisi, 1973; Cipriani et al., 1989; Tonelli, n.d.; Volterra, 1976, 1984) corpus of infant-directed speech. They were obtained from the relevant subcorpora of the CHILDES database (MacWhinney, 2000) by extracting the infant-directed adult utterances. The properties of each corpus are described separately below.

The Hungarian corpus

The CHILDES database contains two Hungarian subcorpora. The MacWhinney corpus (MacWhinney, 1974, 1975) contains orthographic transcripts of recordings of six Hungarian children (age: 1;5–2;10, 3 boys and 3 girls) in their usual kindergarten environment over a period of 10 months. The children interact with each other, the nurses and teachers of the kindergarten, the investigators and occasionally with other adults. The Réger corpus (Réger, 2004) contains orthographic transcripts of the recordings of a Hungarian boy between the ages 1;11 and 2;11 in his family environment. The child interacts with his family members, mostly his mother, occasionally with the investigator and other adults.

The corpus used in the present experiment was derived from the above corpora by extracting all the adult utterances, except those of one investigator, Brian MacWhinney, who is not a native speaker of Hungarian. This way, a corpus of 15 231 utterances, corresponding to 54 881 word tokens and about 8234 word types, was compiled (see Table 3.1). The corpus was purged of material left untranscribed in the CHILDES corpus, but onomatopoeic words, sound imitations, fragments and other linguistic "noise" were kept. All punctuation marks and spaces were deleted, except for utterance boundaries.

The corpus was phonologically transcribed according to the conventions established for Hungarian by the Laboratory of Speech Acoustics of the Budapest University of Technology and Economics within the framework of the BABEL project (Roach et al., 1996). Since Hungarian is a shallow orthography language, first I converted the graphemes into their respective phonemes, then I applied co-articulation and assimilation rules. This procedure, which one may call a "phonotypical" transcription, represents the speech of an idealized educated, middle-class speaker of standard, non-dialectal Hungarian, thus it necessarily abstracts away from potential across- and within speaker variation.

Subsequently, the phonotypically transcribed corpus was syllabified. Syllabification in (adult) Hungarian is unambiguous and follows algorithmically implementable rules (Kiefer, 1994): (i) all syllables must contain exactly one vowel, (ii) a single intervocalic consonant goes into the onset of the second syllable, (ii) two intervocalic consonants are split between the two syllables, (iii) three intervocalic consonants are divided after the second consonant, the first two going to the first syllable, the third to the second. In addition, it was assumed that the definite article cliticizes onto the following noun[8].

The obtained corpus contains 95 816 syllable tokens, falling into 3081 syllable types (see Table 3.1). The average length of words was calculated to be 1.75 syllables. The Hungarian corpus is, therefore, predominantly not monosyllabic.

The Italian corpus

The Antelmi corpus (Antelmi, n.d.) contains the recordings of one Italian child from the age 2;2 to 3;4. The Calambrone corpus (Cipriani et al., 1989) consists of recordings of 6 normally and 11 atypically developing Italian children. The normally developing children were recorded bimonthly in their home environments for a period of about 1.5 years (typically falling between the ages of 1 to 3 years), while the children with language disorder, 3 of which

[8] The definite article has two allomorphs: *a*, used before nouns starting with a consonant, and *az*, used before nouns starting with a vowel. Since the *a* allomorph is a single V, it constitutes a syllable of its own, and poses no problems for syllabification. The *az* allomorph, on the other hand, syllabifies differently according to whether it is a separate word (az#V...) or a clitic (a#zV...). I have decided to treat if uniformly as a clitic, because in colloquial, everyday language it most often cliticizes onto the noun. To confirm this, a small experiment was run with four native speakers of Hungarian, who had to read passages at three different speeds (slow, medium and fast). At medium and fast speeds, which correspond to normal, everyday language use, about 80–100% of the *az* allomorphs cliticized onto the noun and thus necessarily resyllabified (a#zV...).

	Hungarian	*Full Italian*	*Small Italian*	*Japanese*
# utterances	15 231	51 489	10 473	14 958
# word tokens	54 881	233 137	51 138	47 071
# word types	8234	9538	4525	5205
# syllable tokens	95 816	415 334	91 931	79 030
# syllable types	3081	1703	1162	1343
syllables/word	1.75	1.79	1.79	1.68

Table 3.1: Some descriptive statistics of the infant-directed corpora used in Experiment 4.

were followed longitudinally, 8 cross-sectionally, were audio- and videotaped in their institute. The Rome corpus (Antinucci & Parisi, 1973; Volterra, 1976, 1984) contains the recordings of a single Italian boy between the ages 1;4 and 4;0 in his home environment. The Tonelli corpus (Tonelli, n.d.) includes recordings from three Italian children between the ages 1;5 to 2;1. All corpora are in orthographically transcribed format.

I extracted the adult utterances from the above Italian subcorpora, and obtained an Italian infant-directed corpus, which was processed in exactly the same way as the other two. It was purged of untranscribed and uninterpretable material, but linguistic noise was preserved. Punctuation and spaces were removed, except for utterance boundaries. The originally orthographic corpus was phonologically transcribed by first converting the graphemes into their corresponding phonemes, then applying co-articulation and assimilation rules. Finally, the material was syllabified according to the De Mauro dictionary (De Mauro, 2000). The syllabification was manually checked by a native Italian phonologist and two other native speakers, naive to the purpose of the manipulation.

The full Italian corpus. The full corpus contains 51 489 utterances. It is made up of 233 137 word tokens, falling into 9538 word types. Through syllabification, the corpus was broken down into 415 334 syllable tokens, which correspond to 1703 syllable types (see Table 3.1). The average length of words was 1.79 syllables, thus the corpus was found to be exempt from the monosyllabicity problem of Gambell and Yang (2004).

Note that the full Italian corpus is about 3.5 times larger in terms of the number of utterances, and more than 4 times larger in terms of word tokens than the Hungarian and the Japanese ones.

The small Italian corpus. In order to make certain overall distributional analyses comparable between the three languages, a smaller Italian corpus, roughly corresponding to the size of the other two (in terms of word and syllable tokens) was also created by taking the first 5566 and the last 4906 utterances of the full corpus. This small Italian corpus thus contains 10 473 utterances, 51 138 word tokens, 4525 word types, 91 931 syllable tokens and 1162 syllable types (see Table 3.1). Of course, this matching at the level of the number of tokens does not, and in fact cannot, correct for the inherent differences in the number of word and syllable types between the three languages.

3.2.2 Statistical measures

Transition probabilities

I calculated FWTPs according to equation (3.1), and BWTPs according to equation (3.2), where F(AB) is the frequency of unit AB, F(A) is the frequency of unit A, F(B) is the frequency of unit B.

$$TP(\mathrm{A} \rightarrow \mathrm{B}) = \frac{F(\mathrm{AB})}{F(\mathrm{A})} \tag{3.1}$$

$$TP(\mathrm{B} \rightarrow \mathrm{A}) = \frac{F(\mathrm{AB})}{F(\mathrm{B})} \tag{3.2}$$

Mutual information

I computed MI according to equation (3.3), where P(AB) is the probability of the co-occurrence of A and B[9], P(A) and P(B) are the probabilities of A

[9] Note that for these measures to be fully symmetric, both orders of the two elements are to be taken into account. This is how MI is used, for instance, to compute semantic

and B, respectively. P(AB) is obtained by dividing the absolute frequency of the co-occurrence of A and B by the total number of bisyllables in the corpus. P(A) and P(B) correspond to the relative frequency of A and B, i.e. F(A) and F(B) divided by the total number of syllable tokens in the corpus.

$$MI(\mathrm{AB}) = \log_2 \frac{P(\mathrm{AB})}{P(\mathrm{A})P(\mathrm{B})} \tag{3.3}$$

From the point of view of probability theory, MI has a straightforward interpretation. Its denominator contains the probability of the co-occurrence of A and B, when they are *independent* events. If A and B occurs in the corpus with a frequency/probability P(AB) that approximates P(A)P(B), then A and B are independent events in the corpus. In other words, chance co-occurrences are reflected by MI values close to 0 (the fraction gives 1, the logarithm of which is 0). If A and B co-occur more frequently, i.e. P(AB) is greater than P(A)P(B), then A and B are associated. An association is usually considered meaningful at MI values greater than 3. Strong associations start from around 6 (Church & Hanks, 1989; Stubbs, 1995). If the co-occurrence of A and B is less frequent/probable than predicted by their independent probabilities, then A and B are anti-correlated. This is reflected by MI values lower than -2/-3. These threshold values, however, are not absolute. Studies sometimes use different values, because the interpretation of the strength of association also depends on the nature of the data (Church & Hanks, 1989; Stubbs, 1995).

3.2.3 Segmentation algorithms

When using conditional statistics for segmentation, there are at least two ways in which a decision can be made about where to put boundaries. One possibility, used, for example, by Swingley (2005) and Wolff (1975, 1977) is to define global minima or absolute thresholds, and place boundaries where the measures fall below these thresholds. The other option is to use local

relations in a text (Church & Hanks, 1989; Stubbs, 1995). However, in computational studies of syntax (Church & Hanks, 1989), it is more common to compute MI for the AB order only.

minima, and put boundaries where statistical coherence is the weakest in a local context, as done, for instance, by Gambell and Yang (2004). I tested both options for each conditional probability measure.

Global minima When taking the absolute threshold approach, the question that immediately arises is how to determine the value of the threshold. I have decided to use three different methods. First, if it is true that statistical coherence is larger inside linguistic units than at their boundaries, then, as pointed out by Harris (1955), statistical measures will be at their lowest at utterance boundaries. Therefore, the first choice was to use the highest conditional probability values that can be found at utterance boundaries, i.e. between an utterance boundary symbol and the first syllable of the next utterance, as a threshold. If, as I have just argued, probabilities at utterance boundary are indeed the lowest, it can be expected that setting the threshold this way will underestimate the number of word boundaries. Thus it constitutes a cautious estimate. However, it has the advantage of using a non-arbitrary threshold derived from the input material. Secondly, the distributions of TPs and MIs might themselves provide natural thresholds, e.g. they might have mode(s) in the low ranges that correspond to morpheme or word boundaries. Importantly, if such natural thresholds emerge, the morphological differences between the languages tested might be reflected in the number and/or position of the modes that the distributions exhibit in the different languages. Thirdly, a set of a hundred threshold values were chosen arbitrarily for each distribution by taking their 1^{st}, 2^{nd}, 3^{rd} … 99^{th}, 100^{th} percentile. The exact values are reported in Appendix A, Section A.2.2.

As already mentioned, segmentation using FWTP, BWTP and MI yielded one type of boundary, which then was evaluated for whether it fitted word boundaries or morpheme boundaries better. In contrast, segmentation using FW and BW TPs jointly yielded two types of boundaries (see above), low coherence and intermediate coherence boundaries, which were evaluated against word boundaries (WB) and word internal morpheme boundaries (WIMB), respectively (the evaluation procedure is described in detain in Section 3.2.4).

Local minima The other segmentation algorithm I used is a straightforward application of Gambell and Yang's (2004) local minimum rule. A boundary is placed between AB and CD, if TP(A → B) > TP(B → C) < TP(C → D). The potential advantage of such a strategy is to take into account the context, thus the same pair of syllables might or might not be separated by a boundary depending on where they occur. For example, the syllable pair /in/–/kəm/ spans a boundary in *In come Big Bird and Cookie Monster.*, but not in *income.*

3.2.4 Evaluation criteria

Quantitative evaluation

Two kinds of scores were computed to evaluate the obtained segmentation and morphological analyses. As is customary in computational linguistics, I calculated accuracy (as defined in equation 3.4) and completeness (as defined in equation 3.5) scores for each segmentation.

$$accuracy = \frac{hits}{hits + falsealarms} \tag{3.4}$$

$$completeness = \frac{hits}{hits + misses} \tag{3.5}$$

In the original corpora, morpheme boundaries inside morphologically complex words were not marked. Since one of the questions raised here is whether conditional probabilities signal complex morphology, morpheme boundaries were manually added in the morphologically more complex language, Hungarian. This allowed the evaluation of segmentation in three different ways for Hungarian. First, the boundaries resulting from segmentation were compared against the standard word boundaries, as is typically done in the literature. Second, the boundaries posited by segmentation were compared against word and morpheme boundaries taken together. Word boundaries are at the same time morpheme boundaries, since the boundaries of a word necessarily coincide with the boundaries of its constituent morpheme(s), which, in the case of monomorphemic, uninflected words, is the stem itself. I will refer to this evaluation measure as evaluation against morpheme boundaries or MB, in short. Third, when FW and BW TPs were used in combination and segmentation directly provided word boundaries and word internal morpheme boundaries, these were tested against their respective boundary type in the corpus. In the case of Italian, the orthographic boundaries (interword spaces) of the original transcripts were used. Thus, only word level segmentation could be tested, as in most previous studies.

To recapitulate and clarify, segmentation results were evaluated in three different ways in Hungarian, and in one way in Italian. In the former case, the boundaries posited by the segmentation algorithm were compared with (i) the boundaries of morphologically complex word forms (henceforth, word boundaries or WB), (ii) all morpheme boundaries (MB), including both word boundaries (WB) and word internal morpheme boundaries (WIMB), i.e. MB=WB+WIMB, and (iii) word boundaries and word internal morpheme boundaries separately. To give an English example, in the morphologically complex word [*#re-eval.uate-d#*], hash marks indicate WBs, hyphens indicate WIMBs. MBs comprise WBs and WIMBs together. A stem internal syllable transition is shown by the dot. Evaluating segmentation against WBs is the standard procedure in the literature, and for Italian, this is the only evaluation that has been carried out (since the corpus was not morphologically tagged and WIMBs were not available). The effects of agglutination can be estimated by (i) comparing Hungarian and Italian (on WBs), and (ii) by comparing the three different evaluation criteria within Hungarian.

Furthermore, to answer the question about bootstrapping morphological type, the distributions of the conditional probability measures were compared in the two languages.

Qualitative evaluation

The primary aim of this analysis was to assess the sensitivity of segmentation to morphological regularities, especially in Hungarian. For this purpose, certain key features of the Hungarian and Italian morphological systems were chosen as litmus tests. These include the comparison of (i) derivational vs.

inflectional morphology, (ii) verbal vs. nominal inflection, (iii) vowel harmonizing vs. non harmonizing suffixes, (iv) prefixes, suffixes and circumfixes, and (v) allomorphic variants of stems and suffixes.

The segmentation of ambiguous syllable transitions, which contain a boundary in certain contexts, but are stem-internal in others (recall *income tax* vs. *In come John and Mary*), was given special attention. There are three frequent syllables in Hungarian, *–ni* and *–ka/–ke* that can be suffixes (the infinitive marker and the diminutive, respectively), but they can also appear as the last syllables of nouns (*zokni* 'socks'; *macska* 'cat', *kocka* 'cube, die', *szürke* 'grey', *kecske* 'goat'; all examples were taken from the corpus). These were the main targets of the ambiguity analysis. Obviously, segmentation using absolute thresholds cannot handle these cases, since a syllable pair has one conditional probability value assigned to it irrespective of its context. But relative thresholding algorithm may show a better performance.

3.2.5 Results

First, I report the results of the 'global minima at utterance boundaries' algorithm. Second, I present the results obtained when deriving global minima from the distributions of the four conditional probability measures. Third, the segmentation using arbitrary global minima will be discussed. Fourth, results of the local minima algorithm will be reported. For each segmentation algorithm, results will be discussed comparing the two languages and the four measures.

Global minima at utterance boundaries

Below, I will report quantitative segmentation results for the four statistical measures separately. Since segmentation based on global minima at utterance boundaries is one type of absolute threshold, qualitative results will be presented in Section 3.2.5, where a large set of absolute thresholds are evaluated.

FWTP The distribution of FWTP values between pairs of syllables spanning an utterance boundary are shown in Figures 3.1 and 3.2 for Hungarian and Italian, respectively (for exact numerical values, see below). As expected, both distributions are at the lowest range of the TP scale. These values are computed as the frequency of the utterance boundary symbol and the first syllable of the next utterance as a pair, divided by the frequency of the utterance boundary symbol. Therefore, what they ultimately reflect is the relative frequency of syllables at utterance-initial positions (since the denominator is always the utterance boundary symbol). The more frequently a syllable appears utterance-initially, the higher the FWTP value. Only a few syllables show FWTP values that are higher than the lowest extreme. In other words, only a few syllables start utterances with at least some frequency.

Hungarian The majority (84%) of the FWTP values fall below 0.001. The highest FWTP at utterance boundaries is 0.064. This value was chosen as the global minimum threshold.

When compared against word boundaries, this segmentation performed at 72% accuracy and 92% completeness. When tested against word and morpheme boundaries together, an accuracy of 75% and a completeness of 89% were achieved. This can be considered a successful segmentation (cf. previous results presented above, Brent & Cartwright, 1996; Batchelder, 2002; Gambell & Yang, 2004; Swingley, 2005).

Italian (Small corpus) The majority (86%) of the FWTP values fall below 0.001, i.e. most syllables appear once or very few times utterance-initially. The highest FWTP is 0.122. This value was chosen as the global minimum threshold.

The highest FWTP value used as a threshold resulted in 85% accuracy and 55% completeness, which is, once again, a relatively successful segmentation.

BWTP BWTP values at utterance boundaries show a radically different picture. The distributions are illustrated in Figures 3.3–3.4. Unlike FWTPs,

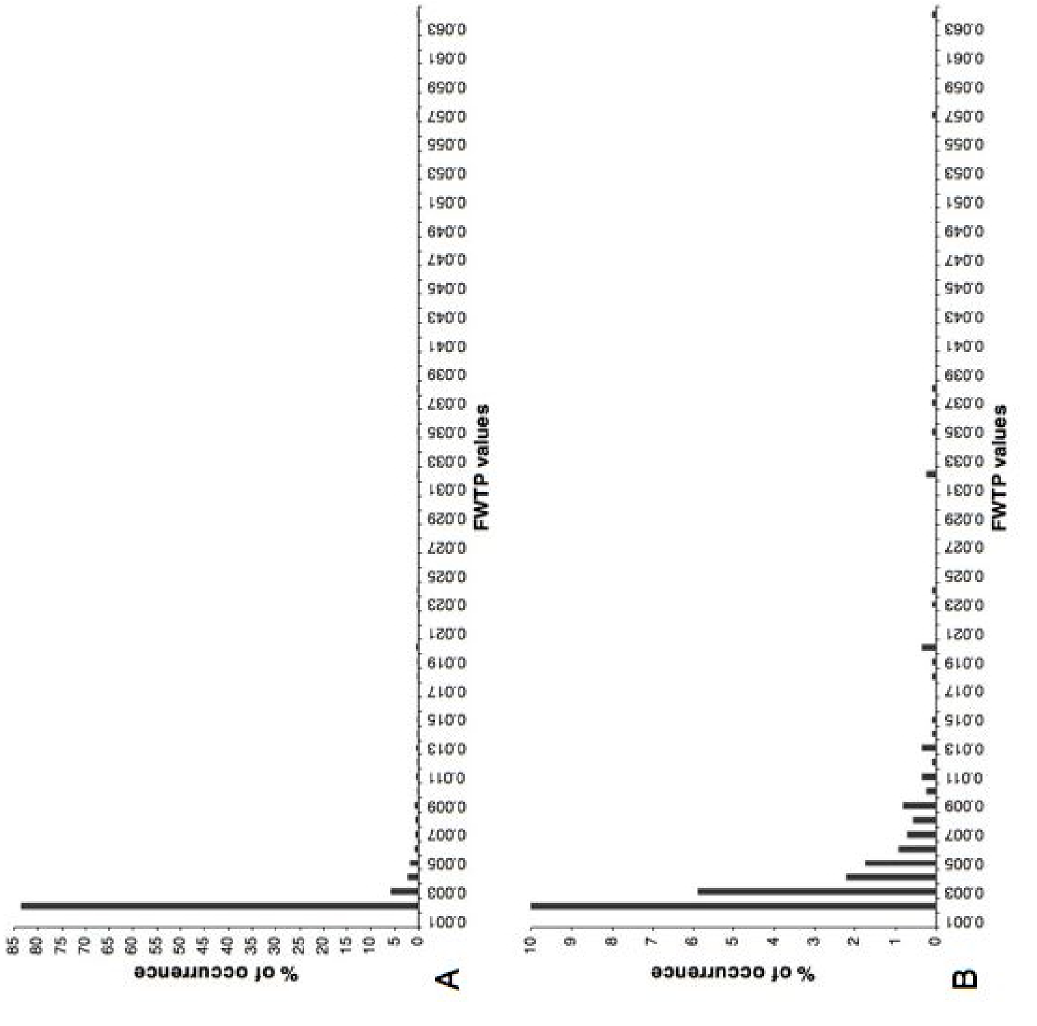


Figure 3.1: The histogram of FWTP values at utterance boundaries in the Hungarian corpus. A: All results. The x-axis represents TP values, the y-axis shows percentage of occurrence, which was chosen instead of absolute frequency to ensure comparability across corpora. The x-axis is plotted only up to the highest TP value found at utterance boundaries, 0.064. B: A zoom on the y-axis. The scale on the y-axis is magnified (cutting off at 10 as the maximum value) for a better visualization of the lower frequency values.

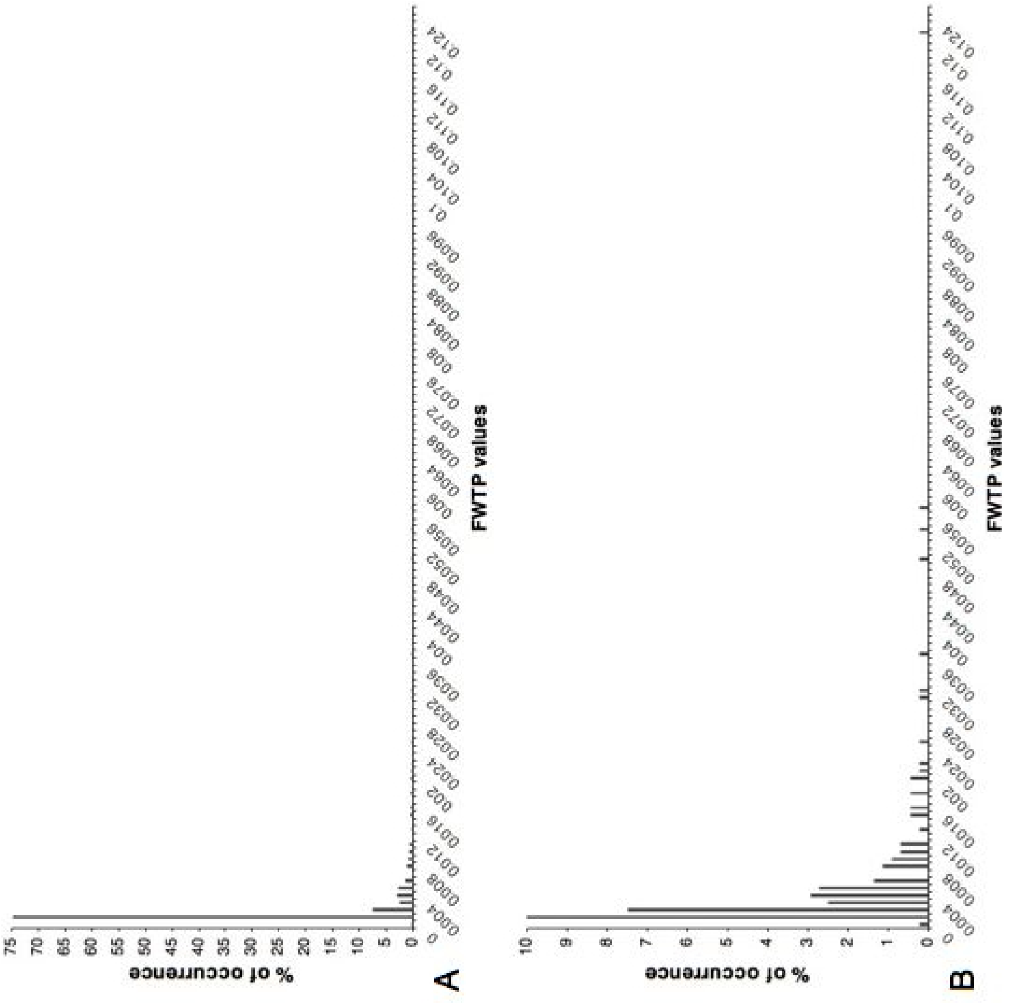


Figure 3.2: The histogram of FWTP values at utterance boundaries in the Italian corpus. A: All results. The x-axis is plotted only up to the highest TP value found at utterance boundaries, 0.122. B: A zoom on the y-axis. All graphical and plotting conventions are identical to those of Figure 3.1.

BWTPs span the whole TP scale from 0 to 1. Moreover, the frequent values are in the high range. However, even the most frequent value, 1, accounts for only a small percentage of the distribution (13% in Hungarian, 7% in Italian). This pattern of results obtains because BWTPs at utterance boundaries are calculated by dividing the frequency of the pair formed by the utterance boundary symbol and the next syllable by the frequency of the syllable. Unlike the highly frequent utterance boundary symbol appearing in the denominator of FWTPs, utterance-initial syllables may be frequent or infrequent. Indeed, as the frequency distributions and the distribution of the FWTPs has already shown, most syllables are infrequent, and even the frequent ones remain less frequent than the utterance boundary symbol. Consequently, in the case of BWTPs, the denominator is a small number. Thus, the resulting TP values will be relatively large. For those syllables that appear only once in the whole corpus and this one occurrence happens to be at beginning of an utterance, the TP value will be exactly 1. In sum, the pattern observed for BWTPs at utterance boundaries reflects the effects of sparse data.

Given this distribution, no value arises as a natural threshold for segmentation. The distribution spans the whole range, and offers no suitable threshold value. Thus, no segmentation was performed for BWTPs.

FWTP and BWTP combined Since no threshold could be obtained for BWTPs, segmentation using the joint values of FWTPs and BWTPs could not be performed.

MI The distributions of MI values at utterance boundaries are shown in Figures 3.5 and 3.6 for Hungarian and Italian, respectively. As with FWTPs, the distributions have clear maxima, which could be used as a threshold for segmentation. In general, the values are within the chance range (between -3 and 3), which indicates that utterance boundaries are indeed points of low coherence.

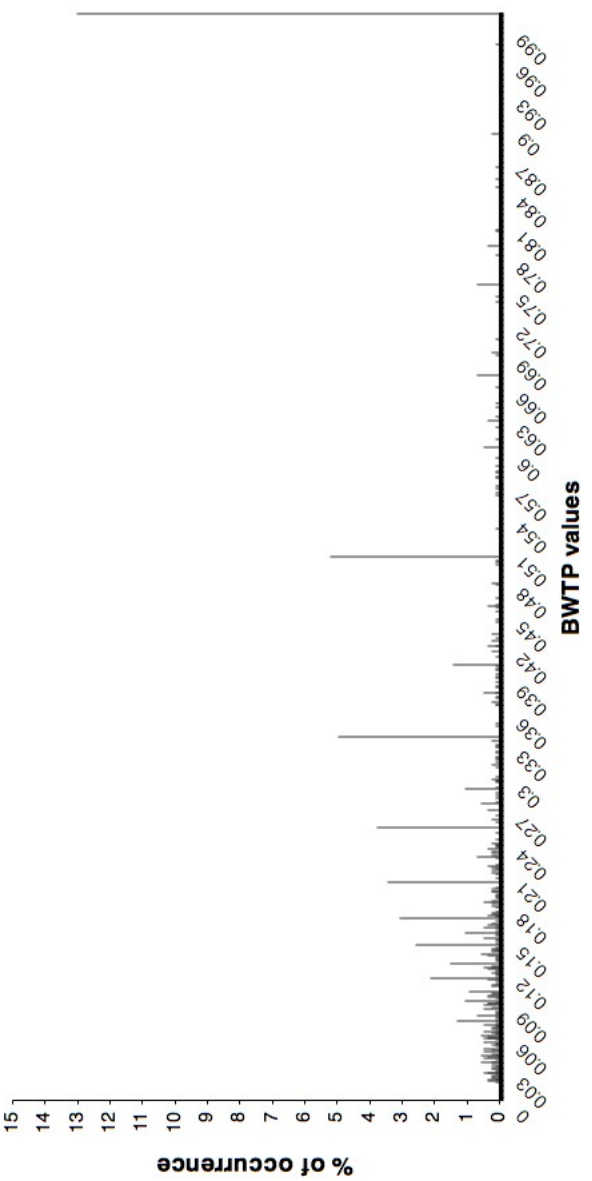


Figure 3.3: The histogram of BWTP values at utterance boundaries in the Hungarian corpus. Graphical and plotting conventions are identical to Figure 3.1.

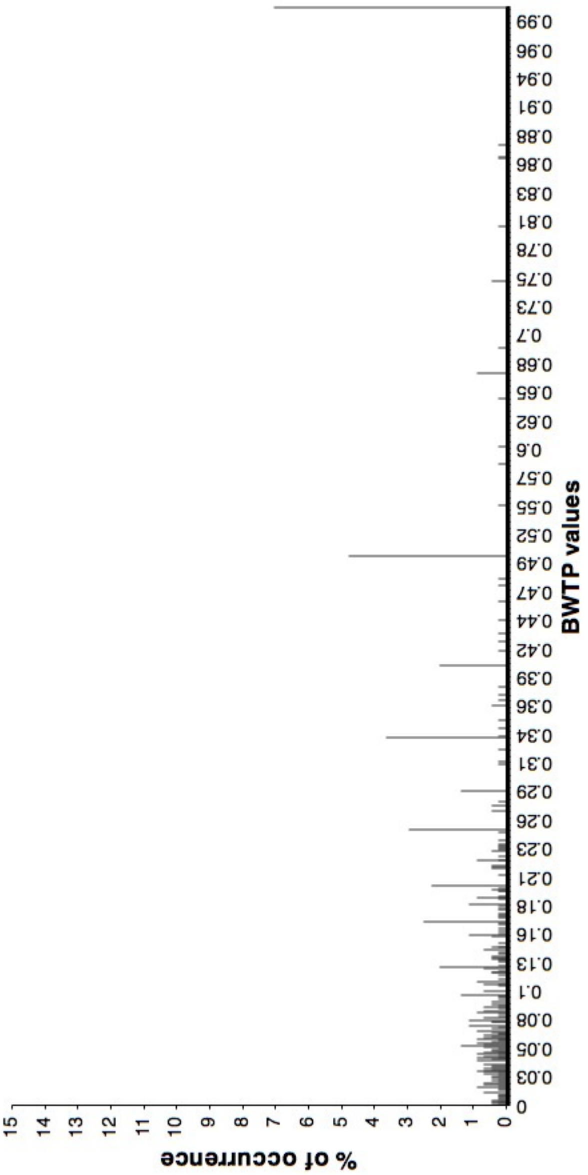


Figure 3.4: The histogram of BWTP values at utterance boundaries in the Italian corpus. Graphical and plotting conventions are identical to Figure 3.1.

Hungarian The highest value, which was used as the segmentation threshold, was 2.9. When compared against word boundaries, the resulting segmentation achieved 90% accuracy and 57% completeness. When morpheme boundaries were considered, accuracy remained 90%, completeness lowered somewhat to 50%.

Italian (Small corpus) The highest value of the MI distribution at utterance boundaries was 3.3. This yielded 70% accuracy and 85% completeness.

Discussion Using the highest conditional probability value found at utterance boundaries, successful segmentation was achieved both in Hungarian and Italian with FWTPs and MI. Segmentation based on BWTPs could not be performed, because values spanned the whole range from 0 to 1. As a consequence, FWTPs+BWTPs could not be used either.

A first conclusion concerning the obtained results is that, contrary to expectations, the segmentation algorithm turned out not to be very conservative. The obtained completeness scores were not particularly low, ranging from 50% to 92%, especially as compared with some previous studies (Gambell & Yang, 2004; Yang, 2004; Swingley, 2005). In other words, this segmentation method did not underestimate word boundaries. This indicates that the coherence between words within an utterance is often not higher than between utterances. The relatively low coherence of the corpus is a result that other segmentation methods will also highlight.

Yet, not all accuracy and completeness scores were equally high, which leads to a second point. An interesting interaction can be observed between the languages and the statistical measures. While FWTPs resulted in lower accuracy and higher completeness scores in Hungarian, but higher accuracy and lower completeness in Italian, MI showed the opposite pattern. The possible reasons for such a difference will be discussed below in Section 3.2.5 together with other similar observations. What is interesting to note at this point is that there is a trade-off between accuracy and completeness, as is often the case in computational studies of language (Charniak, 1993).

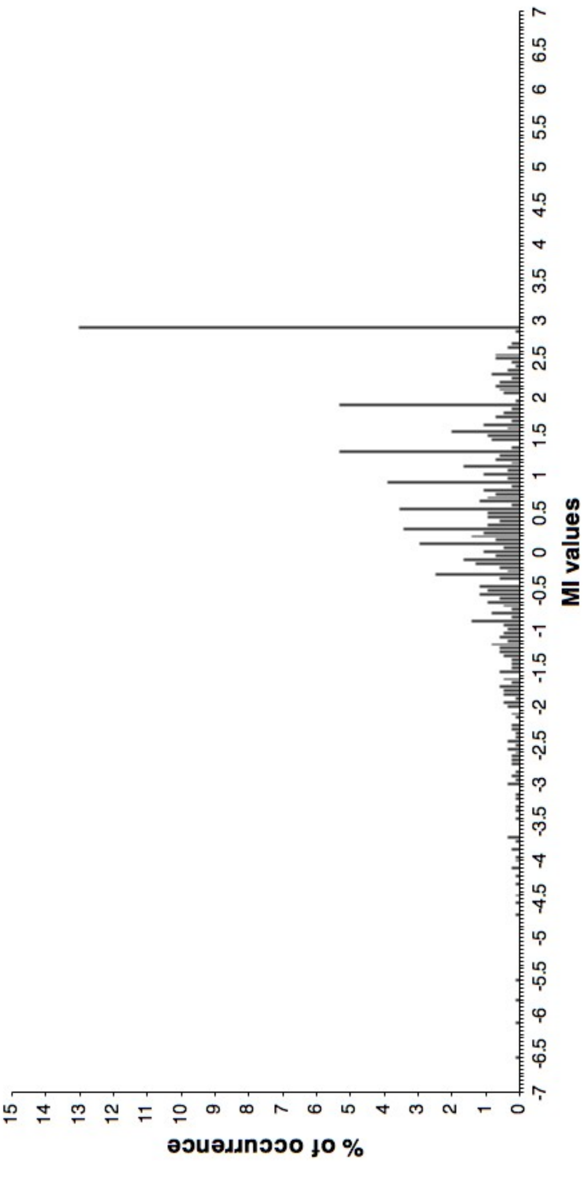


Figure 3.5: The histogram of MI values at utterance boundaries in the Hungarian corpus. The x-axis shows MI values. The y-axis represents frequency of occurrence in percentages.

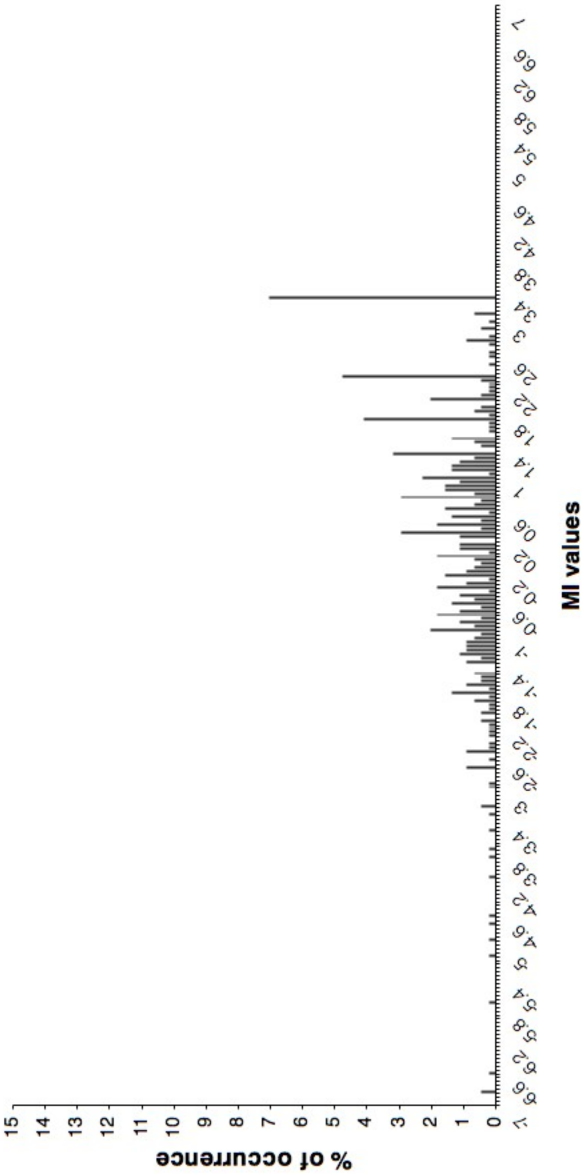


Figure 3.6: The histogram of MI values at utterance boundaries in the Italian corpus. The x-axis shows MI values. The y-axis represents frequency of occurrence in percentages.

This is because these scores are closely related to the number of boundaries posited. If this number is underestimated, accuracy will be high, because the number of hits is high and there are no false alarms (provided the predictions are correct, of course), but completeness will necessarily be low due to the large number of misses, i.e undetected boundaries. This is what I called 'conservativeness' above. If the number of boundaries are overestimated, then no boundary will be missed, so completeness scores are high, but the large number of false alarms will results in lowered accuracy. These two 'strategies' of segmentation will be further discussed in connection with the other segmentation algorithms in Sections 3.2.5 and 3.2.6 below.

A third issue worth noting here is that the highest FWTP and MI values at utterance boundaries are somewhat higher in Italian than in Hungarian. Italian, unlike Hungarian, is Head-Complement, so many of its functors, i.e. the most frequent elements, are at the left edges of syntactic phrases. This might explain why certain syllables are more frequent utterance-initially, resulting in higher FWTP and MI values. This issue will be explored in detail in Experiments 5 and 6.

Global minima from distributions

As argued before, a distribution with clear modes might provide a natural cue to threshold values. Moreover, given the different morphological properties of the two languages, their distributions might show different patterns (e.g. a different number of minima). In the hope of finding such a cue, the distribution of each statistical measure was computed in the two languages. For Italian, both the full and the small corpora were used in order to estimate the effect of the sparsity of the data on the distribution.

FWTP

Distributions The distributions are plotted in Figures 3.7–3.9 for Hungarian, the small Italian and the full Italian corpora, respectively. All three histograms show a power law or Zipfian distribution (Zipf, 1935, see also Appendix A), irrespective of the language. Informally, a few values, those at the lowest ranges of the distribution, appear a large number of times, and a large number of values appear only a few times. (Appendix A demonstrates that TPs indeed follow a power law function, and discusses some theoretical issues related with this observation.) For the ease of comparison and for a better visualization, the lowest ranges (0–0.15) of the three distributions were replotted together in Figure 3.10.

The only observable difference between the distributions is that the Hungarian corpus has higher 'peaks' at the high ranges than the two Italian corpora. These peaks are caused by syllable pairs that appear infrequently and whose first syllable also appears infrequently. These syllable pairs appear to be statistically strongly coherent. However, since they occur only a few times, i.e. the data is sparse, this coherence is (most often) due to chance, rather than a meaningful relation between the syllables. Since the Hungarian corpus has the lowest token/type ratio, it is not surprising that it is the sparsest. The small Italian corpus, which has a similar number of syllable tokens, contains three times fewer syllable types. Even the full Italian corpus, with its four times larger token size, has fewer syllable types than the Hungarian corpus. (For a demonstration of the effects of sparse data on the distribution, see Appendix A, Section A.2.1.)

Power law distributions are very frequent in the quantitative analysis of natural phenomena, including language (e.g. the frequency of words, Zipf, 1935; Cancho, 2005a, 2005b; Cancho & Sole, 2001). So it is not particularly surprising that the same pattern can be observed for FWTPs. As a consequence, however, there is no mode that could serve as a natural threshold value for segmentation.

Segmentation Since the distributions do not have clear modes that could have been used as thresholds, no segmentation could be performed.

BWTP

Distributions The distributions are illustrated in Figures 3.11–3.13 for Hungarian, the small Italian and the full Italian corpora, respectively.

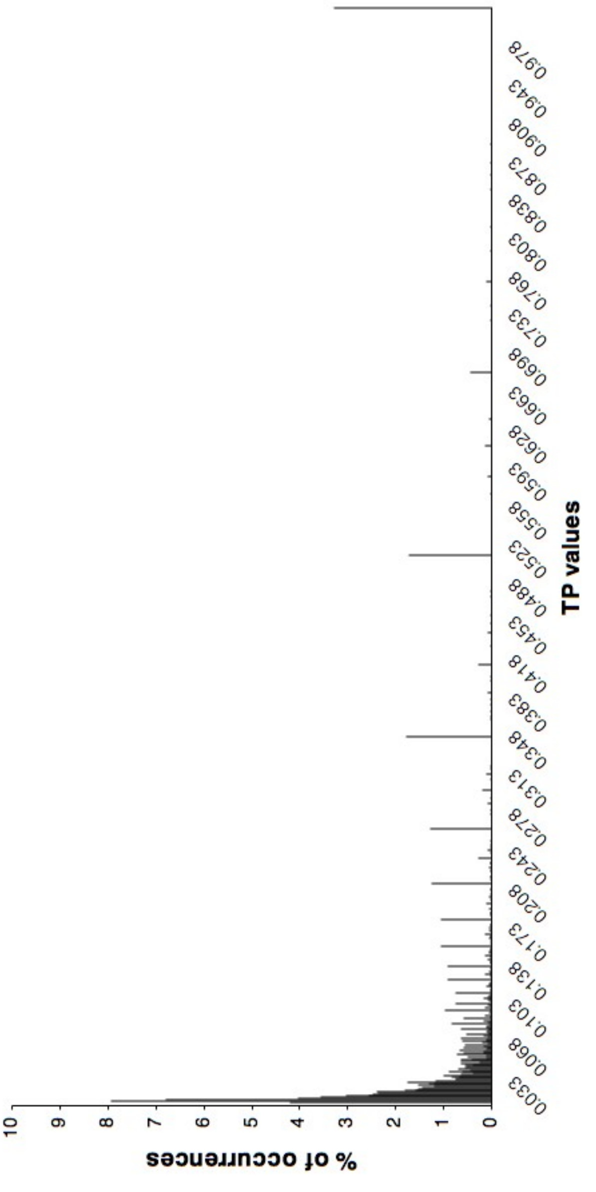


Figure 3.7: Histograms showing the distributions of FWTPs in the Hungarian corpus. The x-axis shows the range of TP values from 0 to 1 in 0.001 increments. The y-axis shows the frequency of occurrence of a given TP value. Frequencies were transformed from absolute values into percentages for cross-linguistic comparability.

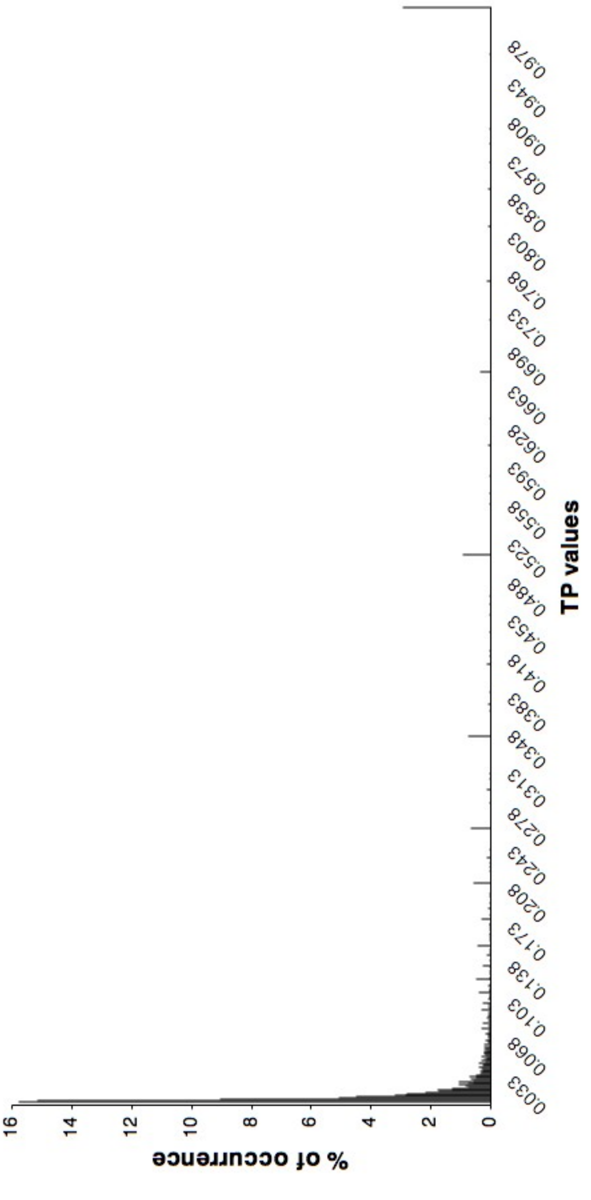


Figure 3.8: Histograms showing the distributions of FWTPs in the small Italian corpus. Graphical and plotting conventions are the same as for Figure 3.7.

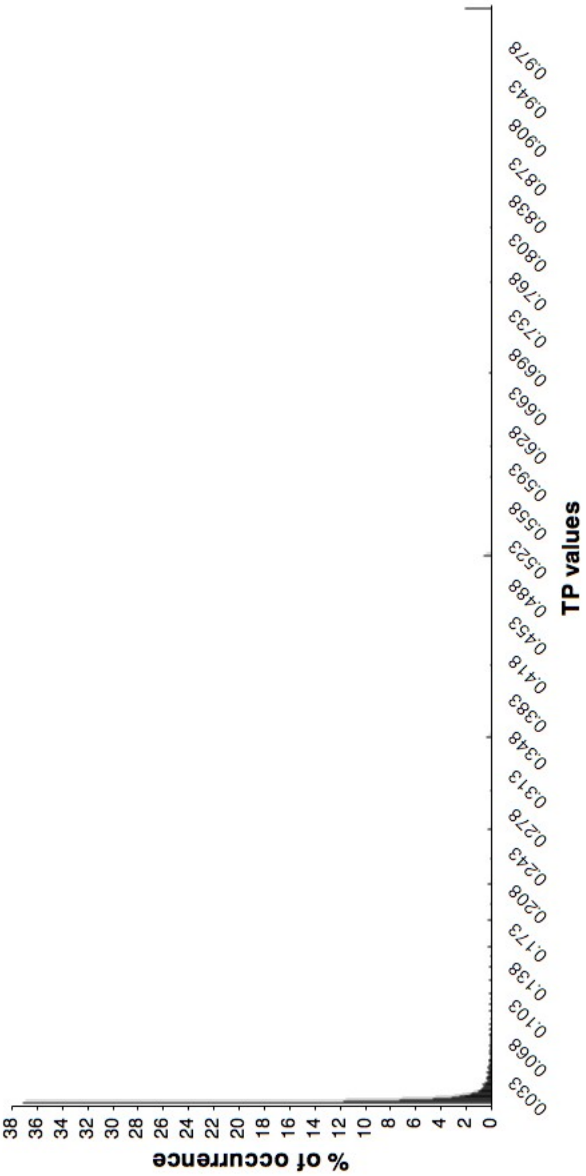


Figure 3.9: Histograms showing the distributions of FWTPs in the full Italian corpus. Graphical and plotting conventions are the same as for Figure 3.7.

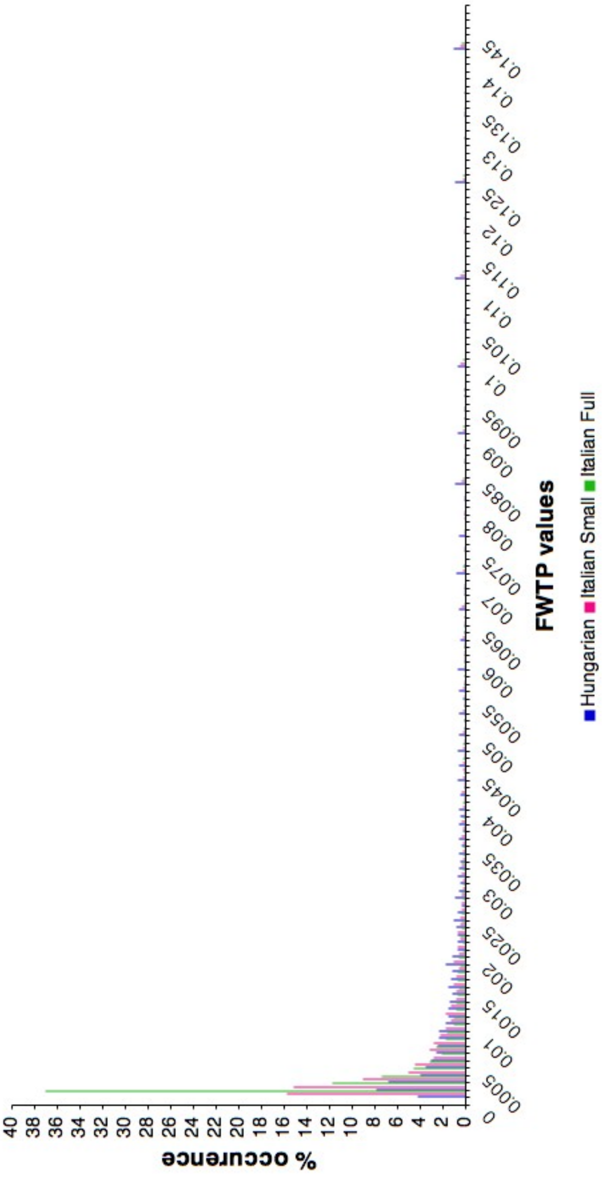


Figure 3.10: Histograms showing the distributions of FWTPs in the Hungarian, small Italian and full size Italian corpora at FWTPs ≥ 0.15. The x-axis shows the range of TP values from 0 to 0.15. All other graphical and plotting conventions are identical to Figure 3.7.

The patterns are very similar to those obtained for FWTPs; values follow a power law distribution in all three corpora. For a better comparison, the 0–0.15 ranges of the distributions are plotted together in Figure 3.14.

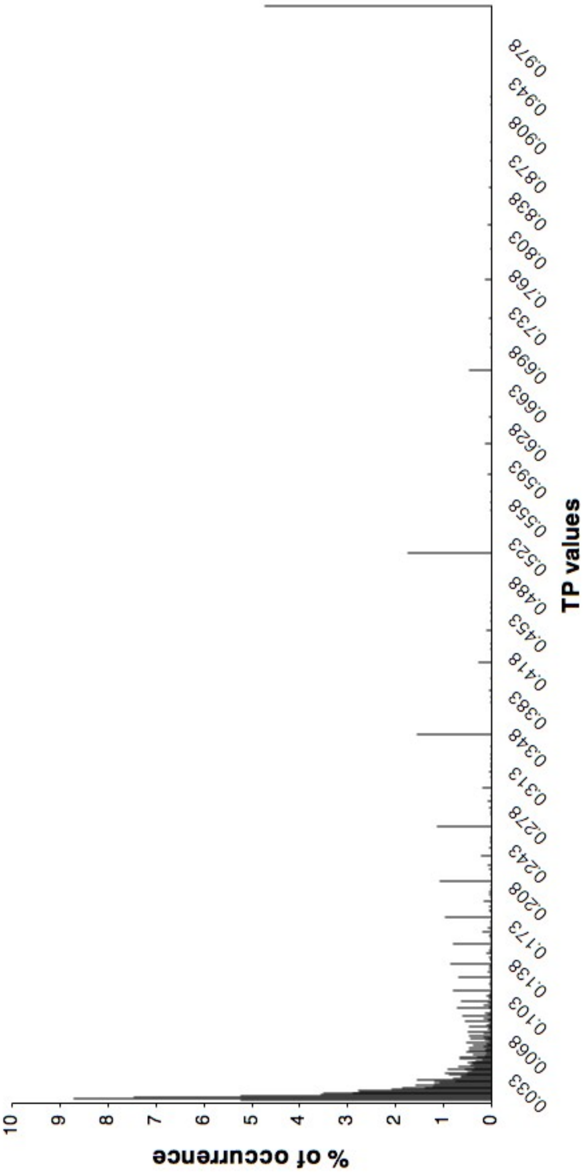


Figure 3.11: Histograms showing the distributions of BWTPs in the Hungarian corpus. Graphical and plotting conventions are the same as for Figure 3.7.

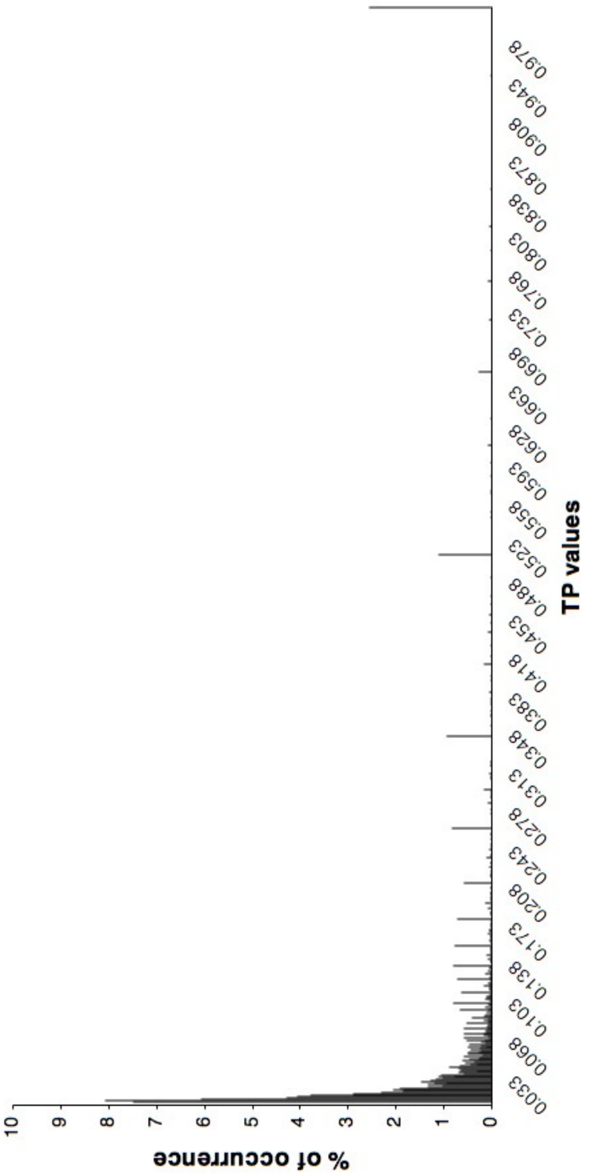


Figure 3.12: Histograms showing the distributions of BWTPs in the small Italian corpus. Graphical and plotting conventions are the same as for Figure 3.7.

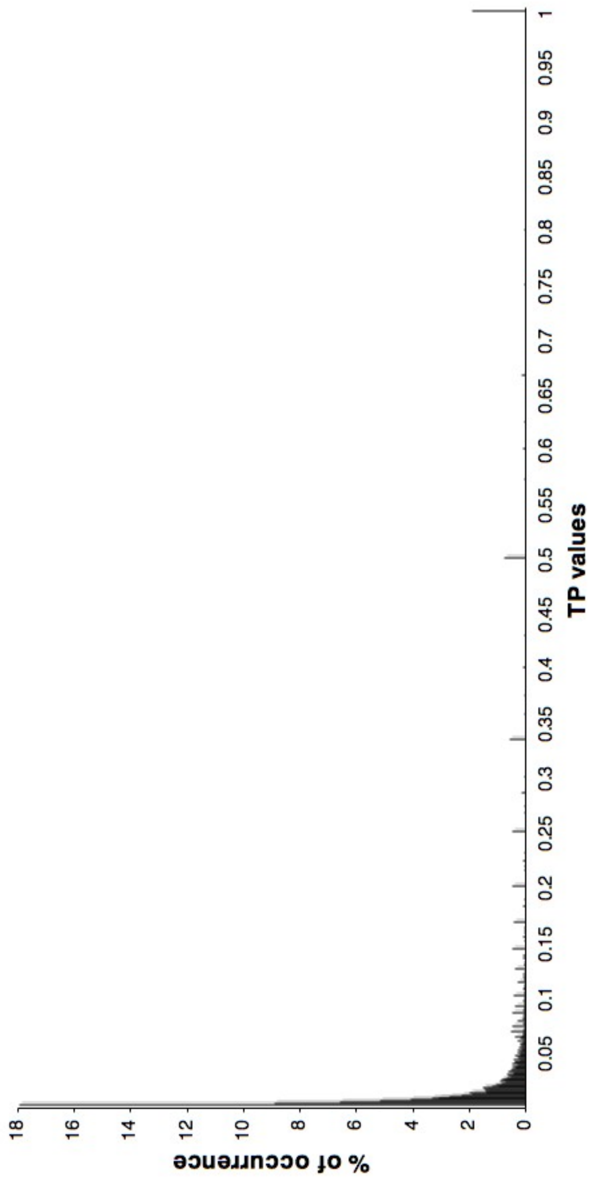


Figure 3.13: Histograms showing the distributions of BWTPs in the full Italian corpus. Graphical and plotting conventions are the same as for Figure 3.7.

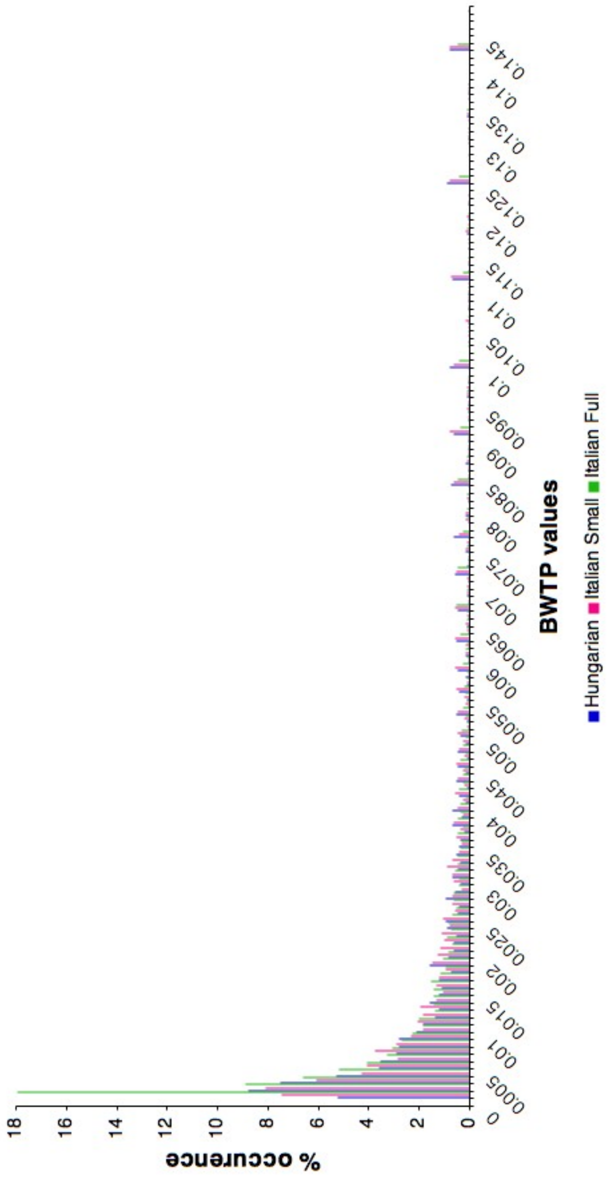


Figure 3.14: Histograms showing the distributions of BWTPs in the Hungarian, small Italian and full size Italian corpora at BWTPs ≥ 0.15. The x-axis shows the range of TP values from 0 to 0.15. All other graphical and plotting conventions are identical to Figure 3.7.

Segmentation Since the distributions do not have clear modes that could have been used as thresholds, no segmentation could be performed.

MI

Distributions The MI distributions for the Hungarian, the small Italian and the full Italian corpora are plotted in Figures 3.15–3.17. The distributions, once again, are very similar in the three corpora (Figure 3.18), they all approximate a positively skewed, platykurtic normal distribution, with occasional peaks corresponding to sparse data. Just as in the case of FWTP distributions, Hungarian shows the most marked signs of sparsity. (For a more detailed discussion on sparsity, see Appendix A, Section A.2.1.)

Most values fall in the chance range, between -3 and 3 in all three languages, showing that a large number of syllable transitions are characterized by low statistical coherence, a property already observed with FWTPs. There are, however, a number of values in the positive tail, which indicate stronger associations. These, just like in the case of FWTPs, mainly represent sparse syllables (see Appendix A, Section A.2.1).

Segmentation Since the distributions do not have clear modes that could have been used as thresholds, no segmentation could be performed.

Discussion The obtained distributions indicate that the transitions between most syllable pairs have a very low probability. In a large corpus, the syllable /prɪ/ typically does not predict the syllable /ti/ with a higher probability than /ti/ predicts /beɪ/, contrary to Saffran, Aslin, and Newport's (1996) famous *pretty baby* example. The distributions of the asymmetric conditional probabilities follow a power law or Zipfian function, the symmetric conditional probability shows a broad and skewed normal distribution. No modes are available as potential threshold values for segmentation.

It has also been shown that the distributions do not differ considerably as a function of the morphological type of the language. Agglutination had no observable effects on the shape and type of the distributions in Hungarian.

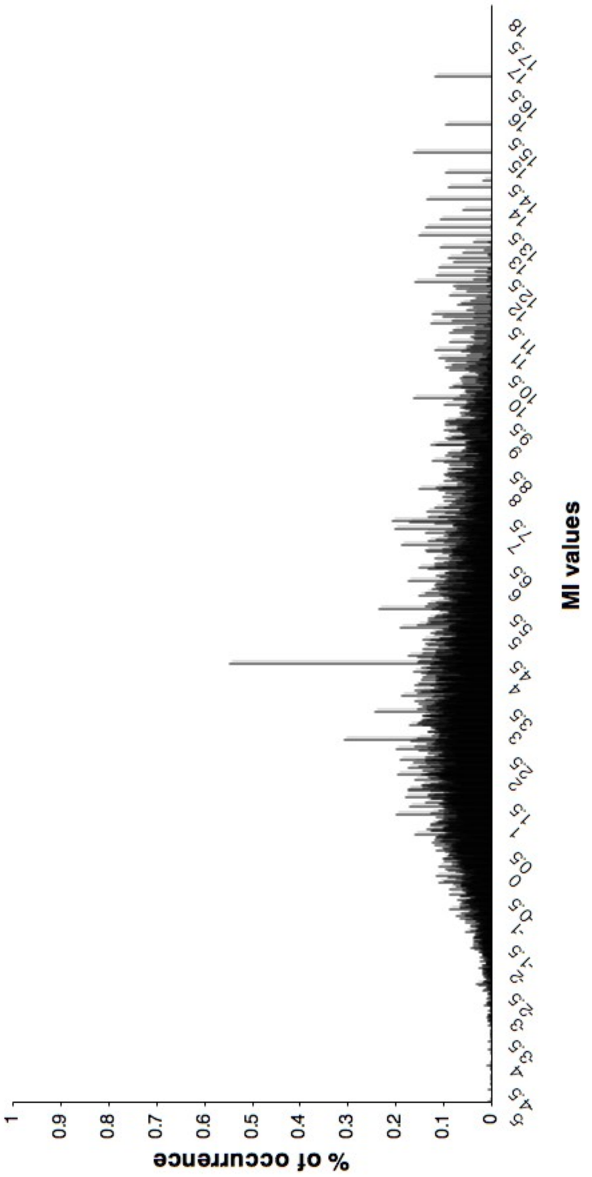


Figure 3.15: Histograms showing the distributions of MI values in the Hungarian corpus. The x-axis shows the range of MI values from -5 to 18 in 0.01 increments. The y-axis indicates the percentage of occurrences for each MI value. Percentages are lower for MI than for TPs, because the value range is divided into a much larger number of increments, so each individual increment has a lower frequency.

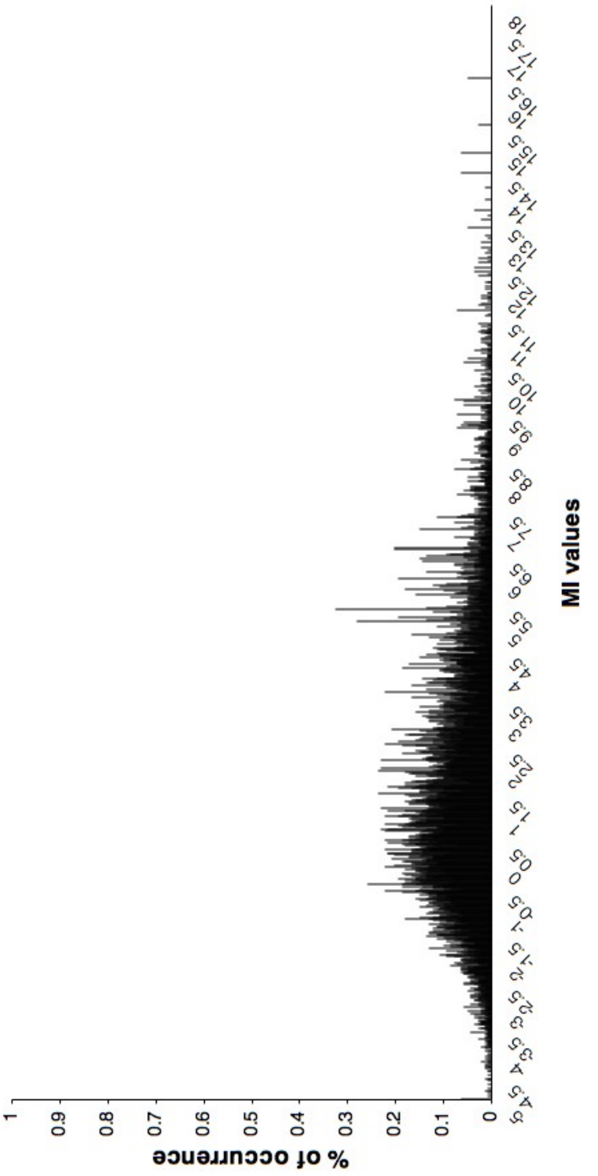


Figure 3.16: Histograms showing the distributions of MI values in the small Italian corpus. Graphical and plotting conventions are the same as for Figure 3.15.

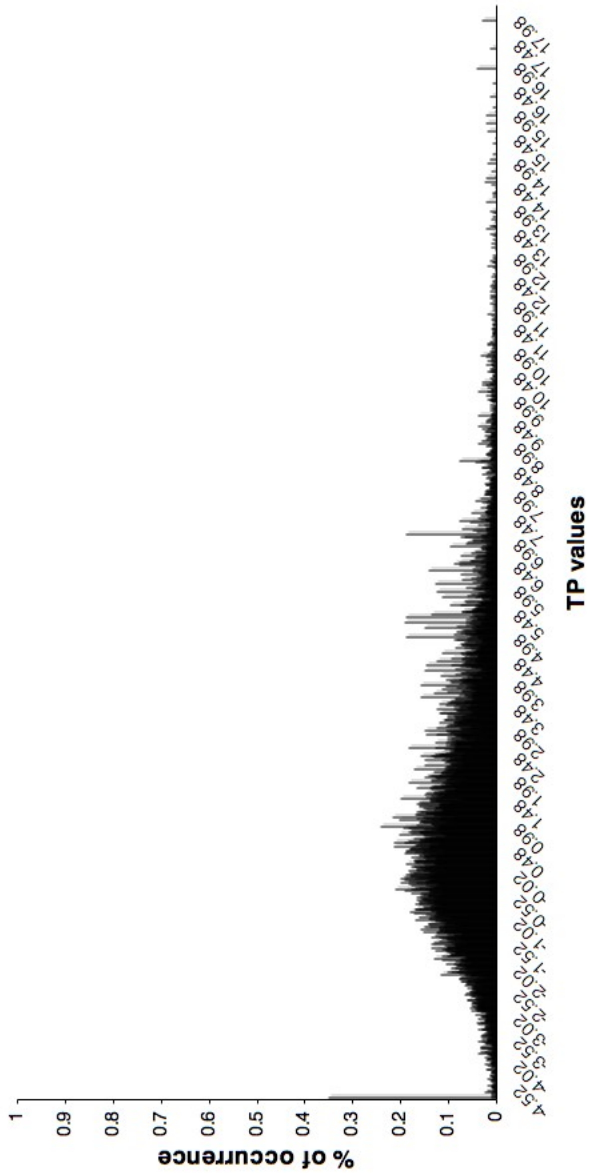


Figure 3.17: Histograms showing the distributions of MI values in the full Italian corpus. Graphical and plotting conventions are the same as for Figure 3.15.

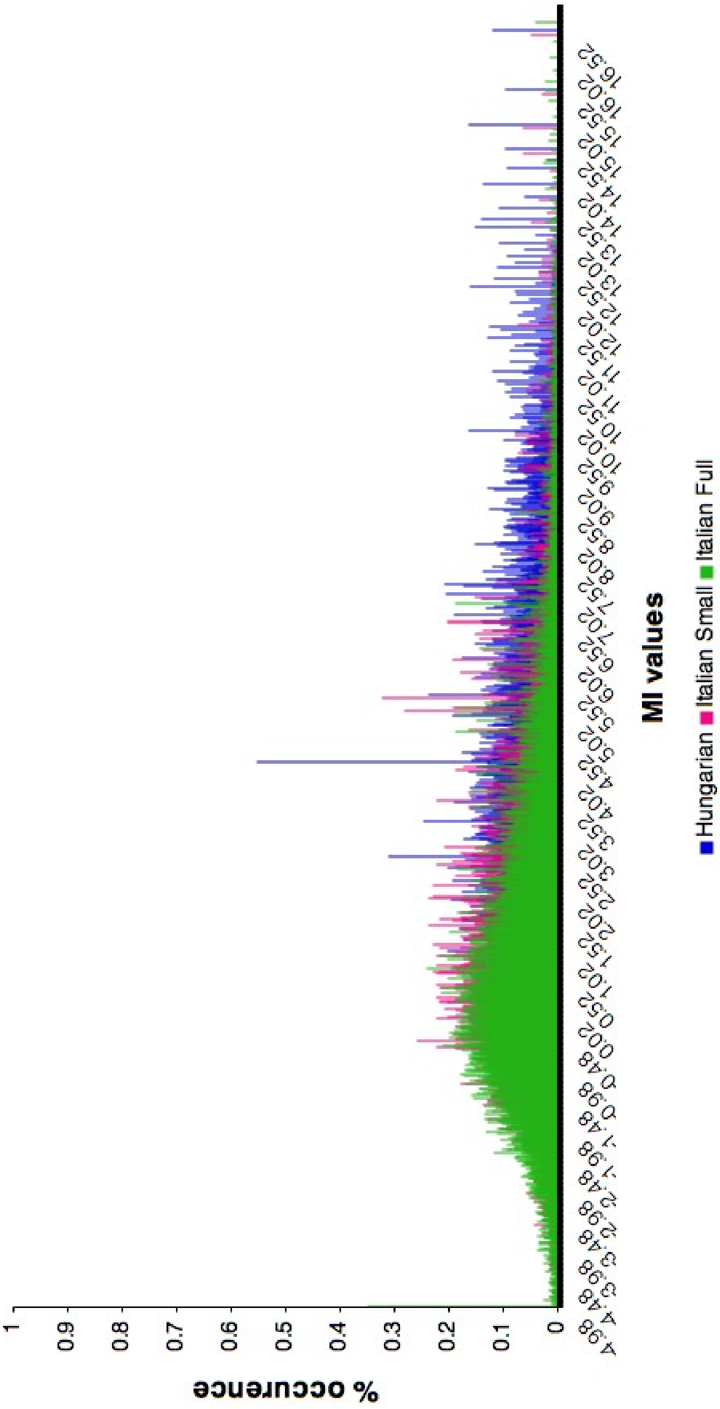


Figure 3.18: Histograms showing the distributions of MI values in the Hungarian, small Italian and full size Italian corpora. Graphical and plotting conventions are identical to Figure 3.15.

The differences that have been found between the corpora are only numerical, and derive from sparsity (see below).

Looking at the findings from another perspective, it can also be argued that what the high frequency of low TP and close to chance MI values reflects is the statistical 'incoherence' of the data. Most syllables appear in the context of most other syllables with some frequency, hence statistically coherent 'islands' emerge relatively rarely, at least when large bodies of data are considered. This has two consequences for statistical segmentation.

First, it might be the case that the main contribution of statistics is the decomposition of the data, while other processes, e.g. prosodic units, ensure cohesion. This possibility will be addressed in some detail on the basis of further segmentation results later.

Second, coherence is higher when data is sparser (cf. a larger number of high TP values in the Hungarian than in the Italian corpora, see also Appendix A, Section A.2.1). The sparsity of the data depends on two factors: corpus size (number of syllable tokens) and the number of different syllable types. If the former decreases, sparsity and thus statistical coherence increases. Consequently, it might be the case that during language acquisition, statistical computations are performed on a small chuck of input at a time. The size of the chunk might be determined by pragmatic considerations (e.g. one person's utterances in a single situation), memory limitations (e.g. the amount of input the infant is able to store at a given time) or linguistic information (e.g. utterance boundaries etc.). The other factor influencing sparsity is the number of syllable types. This, unlike the previous factor, is an inherent property of languages, and is related with syllable complexity. For instance, Hungarian allows complex syllable onsets and codas, while Italian has only the former. This difference is clearly reflected in the number of syllable types found in the three corpora. While it seems to be the case that syllable complexity is correlated with typological properties such as agglutination or word order (Levelt & Vijver, 1998; Mehler et al., 2004; Fenk-Oczlon & Fenk, 2005), it remains an interesting open question whether this relation has any underlying 'cause' in terms of efficiency of encoding in an information theoretical sense, e.g. morphologically more complex languages use more syllable types in order to ensure some level of coherence within words.

Global minima at arbitrary thresholds

Since no ‘natural’ threshold values could be obtained from the distributions, thresholds were arbitrarily chosen using the 1^{st}–100^{th} percentiles.

For Italian, the small corpus was used to ensure comparability with the Hungarian data. Segmentation at the level of morphologically complex word boundaries was performed for both languages. In addition, segmentation at the level of morpheme boundaries (i.e. word boundaries and word internal morpheme boundaries) was also performed for Hungarian.

Hungarian

Quantitative analysis Figures 3.19 and 3.20 illustrate the accuracy and completeness scores, respectively, obtained from the arbitrary thresholds. The figures report both the WB and the MB evaluation criteria.

The results indicate that segmentation is quite accurate with all four statistical measures, especially until about the first 40-50 percentiles, where accuracy remains above 80%. Above the 50^{th} percentile, accuracy starts to drop and reaches its minimum (60%–65%) around the 90^{th} percentile. Although differences are small between the measures, BWTPs and MIs perform better than FWTPs+BWTPs, which in turn outperform FWTPs, particularly in the first 50 percentiles.

Completeness shows the opposite pattern. Scores start at 0 and increase linearly until they reach their maximum at around the 90^{th}–95^{th} percentile. While the scores augment linearly for all four measures, growth is faster for FWTP+BWTP and MI. The advantage of these two measures is most pronounced after the 40^{th}–50^{th} percentiles.

No difference has been found between the evaluation relative to WBs and the evaluation relative to MBs, except in the highest ranges (roughly above the 70^{th} percentile), where segmentation evaluated against MBs was slightly more accurate. This happened, because MBs are a superset of WBs.

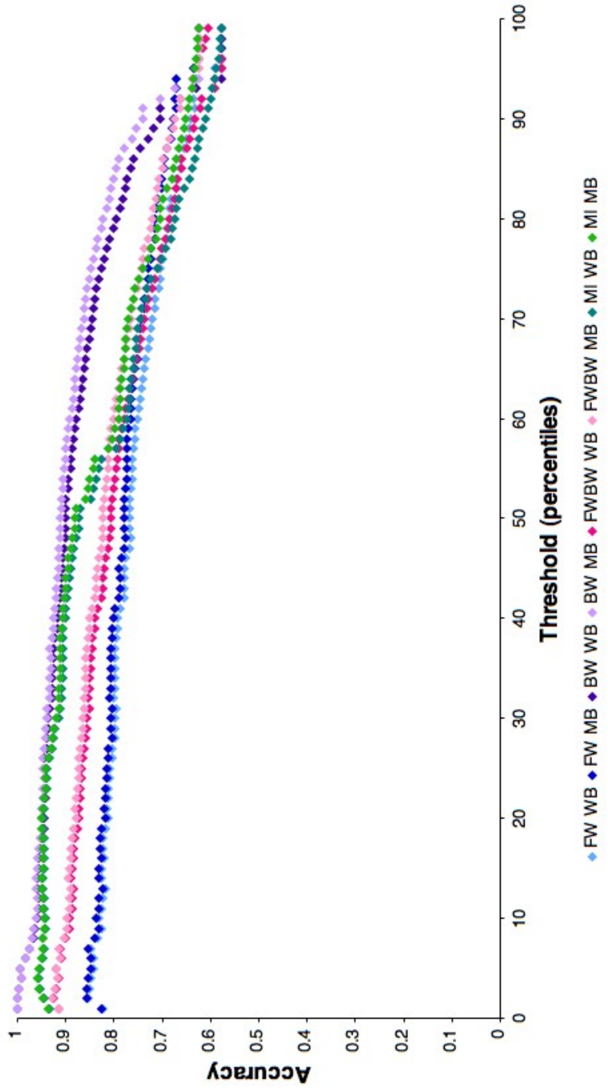


Figure 3.19: The accuracy of the segmentation obtained by using arbitrary thresholds for each statistical measure, evaluated against the WB and MB criteria in the Hungarian corpus. The x-axis shows the thresholds, defined as the 1st–100th percentiles of each distribution. The y-axis represents the accuracy scores.

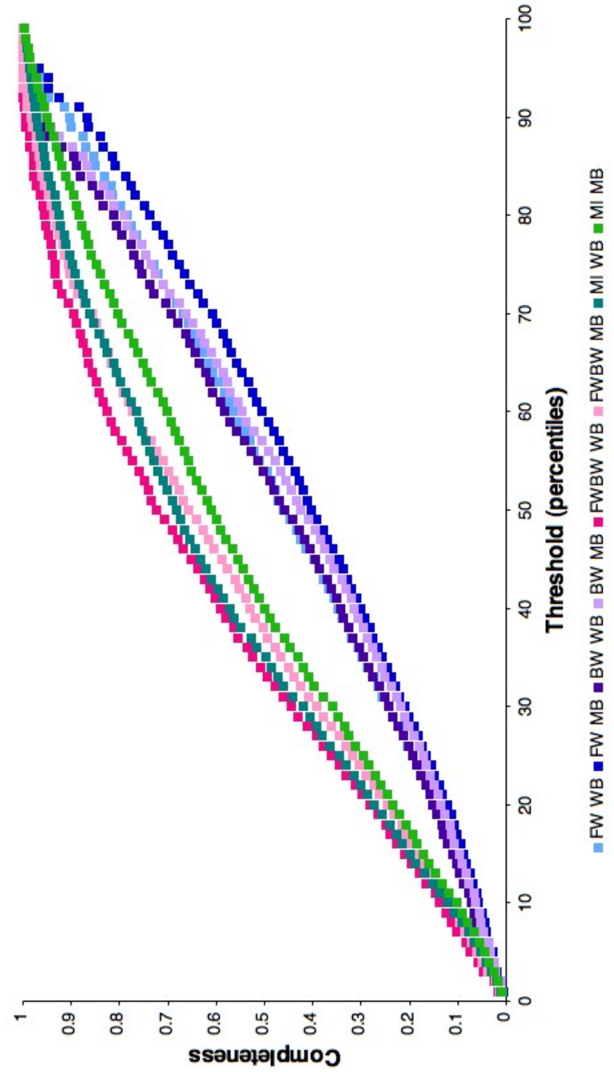


Figure 3.20: The completeness of the segmentation obtained by using arbitrary thresholds for each statistical measure, evaluated against the WB and MB criteria in the Hungarian corpus. Graphical and plotting conventions are identical to Figure 3.19.

Therefore, their number is higher, so they give rise to a higher number of hits.

When the detection of word boundaries and word internal morpheme boundaries were evaluated separately using the combination of FWTPs and BWTPs, WBs were segmented much more accurately, but less completely than WIMBs, as illustrated by Figure 3.21. The reason for the low accuracy of WIMBs was the high number of false alarms. Since the segmentation algorithm posited a morpheme boundary whenever one TP was below the threshold, while the other was above, WIMBs were generated even when the syllable transition was coherent in one direction. This resulted in a considerable overgeneration of WIMBs. For WBs, on the other hand, both TPs needed to be below the threshold, i.e. coherence was required to be very low in both direction, which ensured a conservative prediction of WBs. This resulted in high accuracy, but low completeness.

Qualitative analysis As discussed above, after about the 50th–60th percentile, the accuracy of the segmentation starts to degrade, because boundaries are overgenerated. Therefore, the qualitative analysis will mostly concentrate on the lower percentile ranges, where the qualitative differences between the statistical measures are more marked. Indeed, in the first 50 percentiles, transition probabilities and MI show different patterns of results. FWTPs, BWTPs and their joint application first segmented out the boundaries (left in the case of BWTPs, right in the case of FWTPs, either for FWTPs+BWTPs) of the most frequent syllables. Indeed, in the first percentile, the only word boundary correctly identified is the boundary of the definite article *a*, which is the most frequent syllable in the corpus. After the second percentile, the boundaries of other functors, such as *mi* ‘what’, *ki* ‘who’, *van* ‘is’, *te* ‘you’, *és* ‘and’, *de* ‘but’, are also detected. The correct segmentation of bound morphemes appears in the 6th percentile with the infinitival suffix -*ni* and the preverbal prefix/particle[10] *le* ‘down’. After about

[10] Preverbal particles are not prefixes in the classical sense. They are functional morphemes that are prefixes to the verb or follow it as a free morpheme depending on the syntactic context, e.g. *le-ülsz* ‘down sit.2SG ‘you (are) sit(ting) down’ vs. *ülj le!* sit.2SG.IMPER ‘sit down’. Here, they are uniformly treated as prefixes.

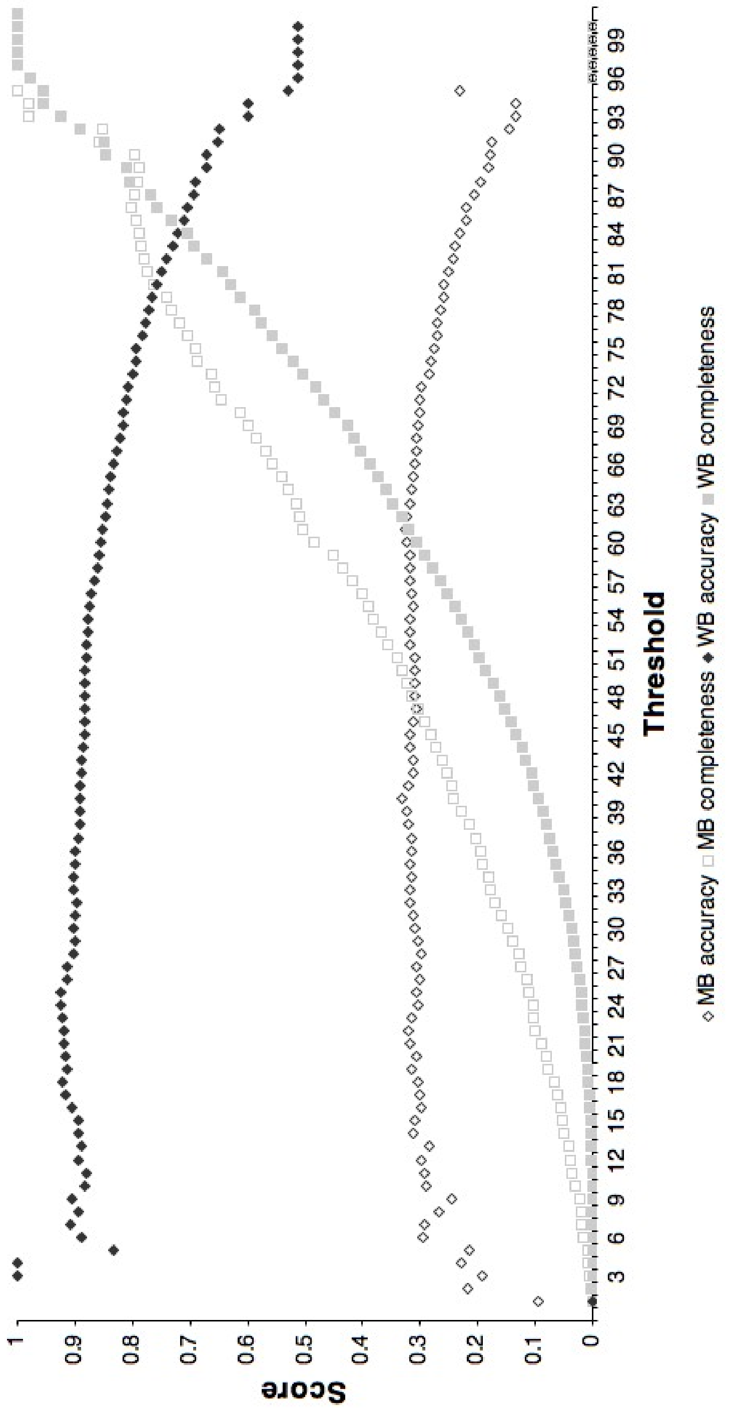


Figure 3.21: The accuracy and completeness of the segmentation obtained by jointly using FWTPs and BWTPs to generate word boundaries and word internal morpheme boundaries. Graphical and plotting conventions are identical to Figure 3.19.

the 8th–10th percentile, a wide variety of verbal and nominal suffixes appear among the correctly segmented items. The suffixes include both harmonizing and non-harmonizing ones. In general, the more frequent a monosyllabic word, the earlier its boundaries are detected. This is especially true of functors, which appear in a wide range of different contexts. It has to be noted, though, that even the most frequent monosyllables fail to get segmented out in some environments. In their most frequent collocations, e.g. *ab-ban* that.INESS 'in that', *szobá-ban* room.INESS 'in the room', they get segmented out only after the 50th–70th, when boundaries are overgenerated.

The close connection between the frequency of a syllable and its good segmentability can be explained by the mathematical definition of TPs. The TP value of a syllable pair is conditioned on the frequency of one of its members (the first for FWTPs, the second for BWTPs). TP values are low, i.e. easily segmentable, if the denominator is a large number, i.e. the conditioning syllable is a lot more frequent than the syllable pair. This is typically the case with functors and a few high frequency content words, which appear in numerous different contexts and are usually not strongly associated with other items. It is interesting in this regard to compare the definite article *a* and the non-harmonizing infinitival suffix *-ni* with the harmonizing inessive suffix *-ban/-ben*. The former get segmented out earlier than the latter (although the difference is not large), because the former come in only one allomorph, so they are less context-selective than the harmonizing forms (while all of them have roughly the same frequency).

Following the above logic, FWTPs are better predictors of boundaries when the first item of a syllable pair is more frequent, whereas BWTPs perform better when the second item is more frequent. Since Hungarian is agglutinating and Complement-Head, segmentation is more successful backwards, since the frequent, thus easily segmentable items come second/last within linguistic units. However, FWTPs also achieve some accuracy, because Hungarian is Spec-Head, so frequent determiners precede their nouns.

Unlike TPs, MI does not segment out clear morphological categories. From the lowest to the highest percentiles, boundaries are posited at points of weak associations, which might be a transition between two frequent syllables, but also a transition between two medium or even low frequency syllables that simply happen to co-occur rarely.

Italian (Small corpus)

Quantitative analysis Figures 3.22 and 3.23 illustrate the accuracy and completeness scores, respectively, obtained from the arbitrary thresholds.

In Italian, the four measures segment the corpus with different accuracy. FWTPs yielded a highly accurate segmentation (around 90%) that declined only after the 65th–70th percentile. BWTPs, in contrast, achieved much lower accuracy (55%–60%), which remained constant across the entire range. The combination of FWTPs and BWTPs produced low accuracy scores at the low and high extremes of the range (below the 20th and above the 70th percentile), but yielded high scores (90%) in between. MI, somewhat like FWTPs, also achieved high accuracy (80%) up to the 70th percentile, after which a decline followed.

Completeness scores started out low and increased linearly for FWTPs, BWTPs and MI, the latter outperforming the other two. The increase was initially slower, but then more abrupt for the combination of FWTPs and BWTPs, the completeness score of which stayed close to 0 almost until the 40th percentile, then plateaued at 1 after the 80th percentile.

Qualitative analysis The Italian segmentation results parallel the Hungarian ones. TPs segment out the boundaries of the most frequent syllables first. In Italian, the first category to be segmented out is the connective *e* ‘and’. The definite article, which was the first in Hungarian, gets segmented out only after the 3rd–5th percentile. This difference between Italian and Hungarian might be related to the fact that the definite article in Italian is more context selective than in Hungarian, since it has seven allomorphs varying according to the gender, the number and the initial consonant (cluster) of the subsequent noun, while in Hungarian, only two allomorphs exist, depending on the initial sound of the noun. This results in a larger reduction of the individual allomorph frequencies in Italian than in Hungarian.

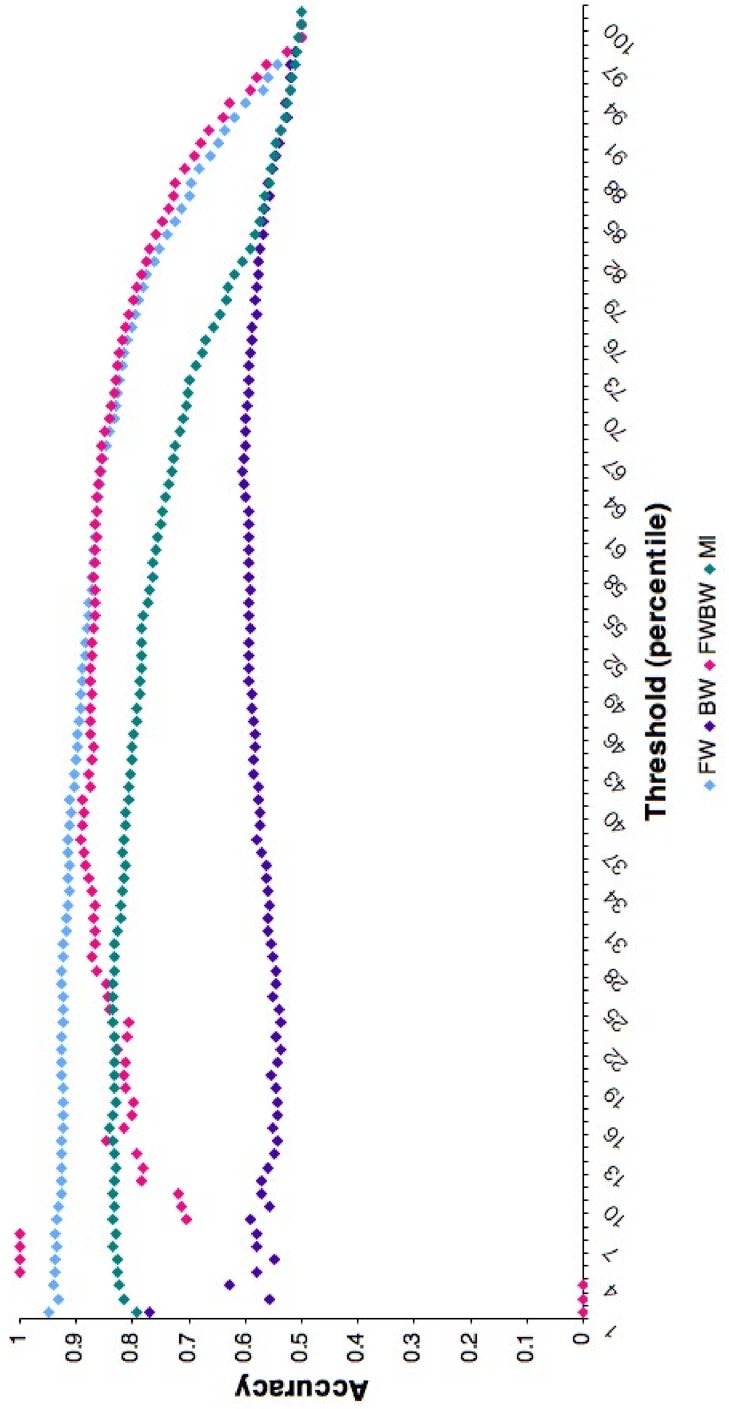


Figure 3.22: The accuracy of the segmentations obtained by using arbitrary thresholds for each statistical measure in the small Italian corpus. Graphical and plotting conventions are identical to Figure 3.19.

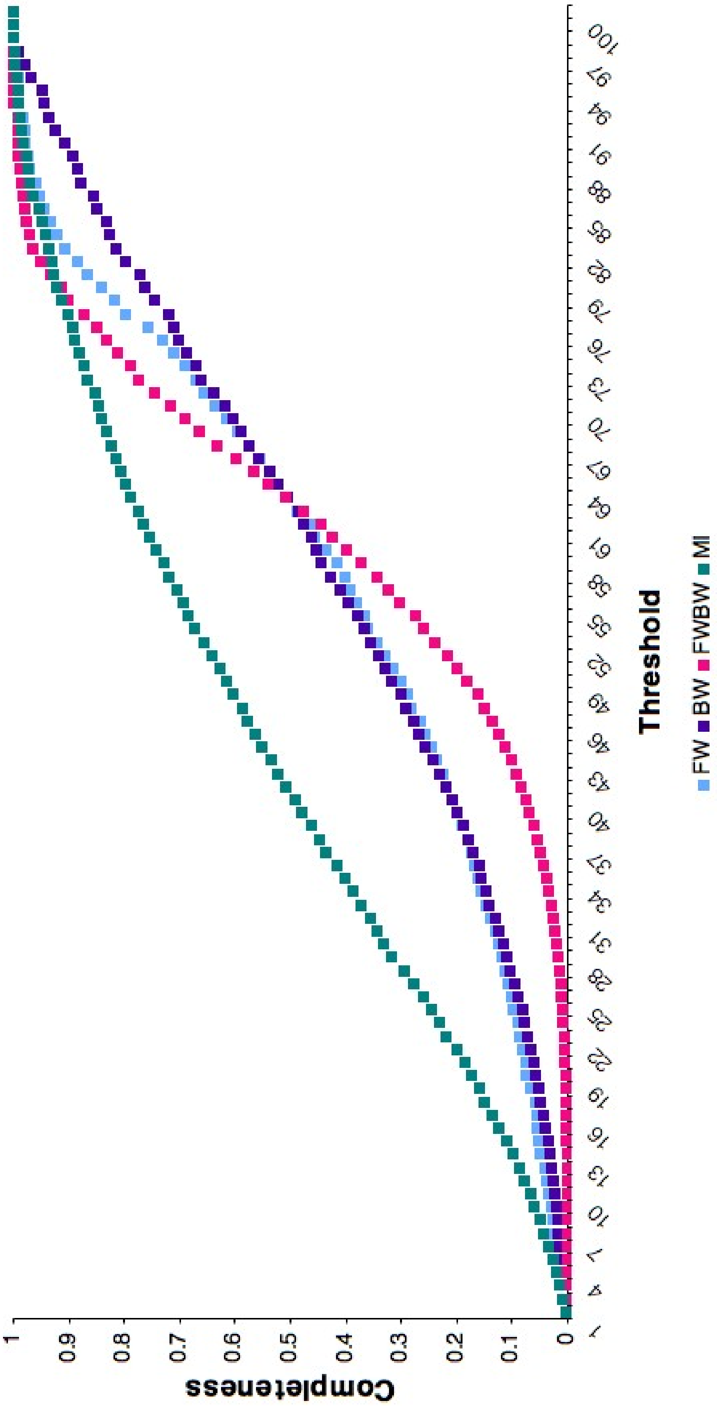


Figure 3.23: The completeness of the segmentations obtained by using arbitrary thresholds for each statistical measure in the small Italian corpus. Graphical and plotting conventions are identical to Figure 3.19.

Prepositions are segmented out from the 10th–15th percentile.

Unlike in Hungarian, however, FWTPs produce more accurate segmentations than BWTPs, and the difference between the performance of the two measures is much greater than in Hungarian. The reason for this is that Italian is not agglutinating, Head-Complement and Specifier-Head. Therefore, all frequent elements appear in initial positions.

Cross-linguistic comparison For a better comparison, the results of the two languages are plotted together in Figures 3.24 and 3.25. (For Hungarian, the results relative to WBs are used to parallel the Italian case.)

Except for the Italian BWTPs, the different measures all give rise to highly accurate (>80%) segmentation. However, systematic differences between the two languages emerge. First and foremost, there is an interaction between the simple asymmetric conditional probabilities and the two languages. While FWTPs are better predictors of segmentation in Italian than in Hungarian, the opposite is true of BWTPs. This difference is larger for the latter measure than for the former. The relations between the bidirectional measures are more complex and depend on the range. In the lowest ranges, FWTPs+BWTPs and MI are more accurate for Hungarian than for Italian. This remains to be the case for MI throughout the entire range, but the difference considerably decreases after the 60th percentile. For FWTPs+BWTPs, in contrast, the Italian accuracy scores increase quite importantly and remain superior to the Hungarian ones almost until the highest extreme of the range.

Completeness scores are also similar in the two languages for almost all measures, increasing from 0 to 1 in a roughly linear fashion. However, Hungarian completeness scores are systematically higher than the respective Italian ones.

Discussion Segmentation based on absolute thresholds has proven to be reasonably reliable in both languages with all four statistical measures. A trade-off has been observed between accuracy and completeness as a function of the percentile range. When thresholds are low, segmentation is usually

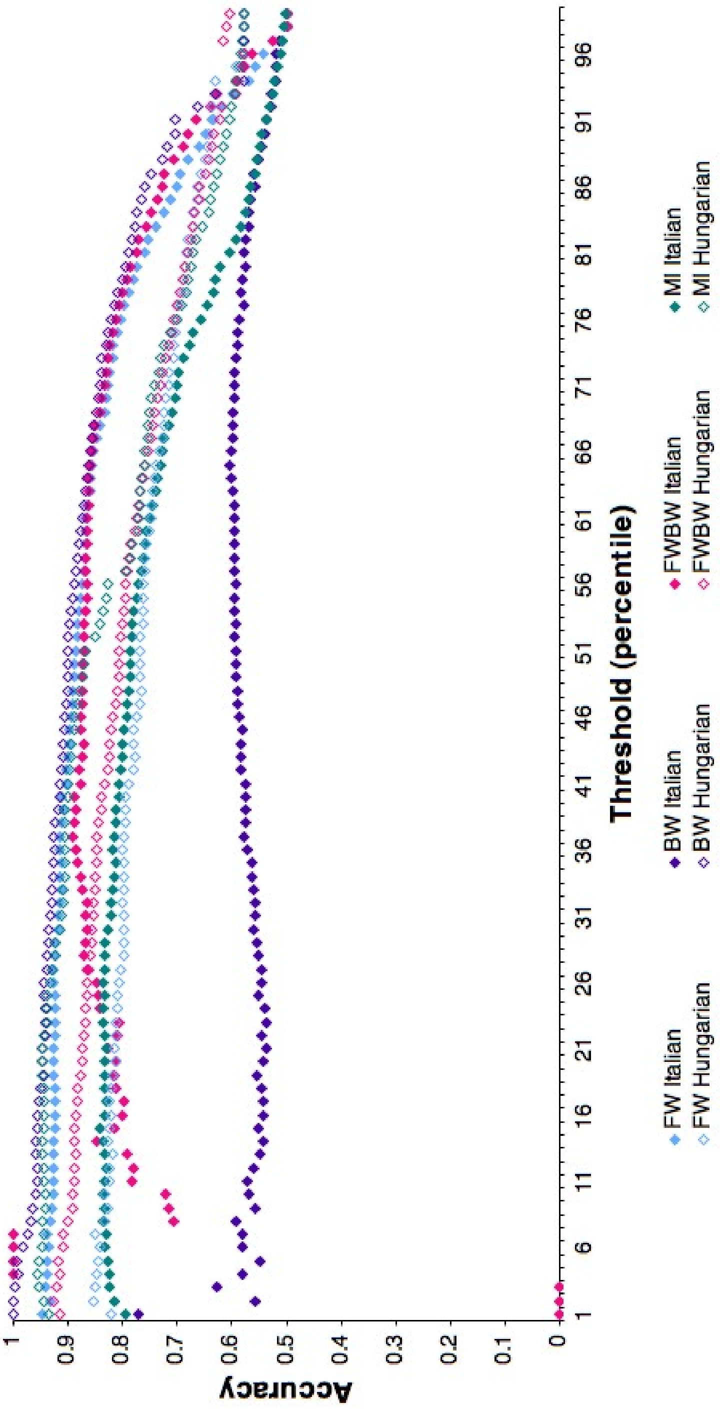


Figure 3.24: The accuracy of the segmentations obtained by using arbitrary thresholds for each statistical measure in the two corpora. Graphical and plotting conventions are identical to Figure 3.19.

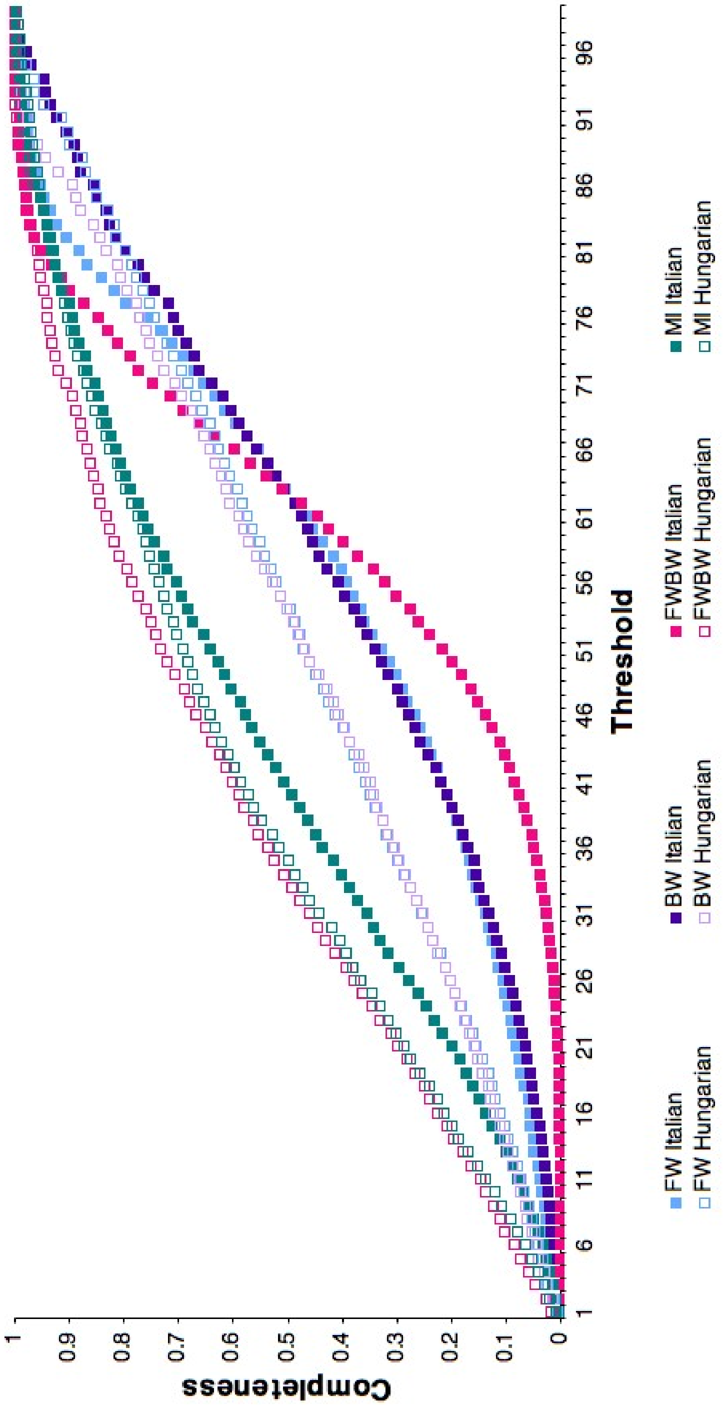


Figure 3.25: The completeness of the segmentations obtained by using arbitrary thresholds for each statistical measure in the two corpora. Graphical and plotting conventions are identical to Figure 3.19.

highly accurate, but fails to detect the majority of the boundaries. At higher thresholds, accuracy decreases, while completeness increases. This trade-off is a direct consequence of the number of boundaries that are predicted. At low thresholds boundaries are undergenerated; at high thresholds, they are overgenerated. Whether this balance between over- and underestimation has any cognitive parallel in the process of language acquisition remains an open question. It might be the case that the preferred strategy changes throughout development. For instance, at the beginning, a conservative estimate might be used to ensure correct segmentation, but only a few words are learned. Once many words are already known (and maybe other language-specific segmentation cues, such as phonotactics or stress patterns, have also been acquired), a larger number of boundaries can be posited.

One important finding of all the global minima algorithms is that different statistical measures perform with different accuracy in the two languages. FWTPs provide better segmentation in Italian, while BWTPs are more efficient in Hungarian. This difference is due to the morphological and syntactic properties of the two languages. Consequently, the predictive value of the statistical measures provides a possible bootstrapping cue. How infants might 'know' about the accuracy of a statistical measure is a question that future research will need to address. They might use the measure's correlation with other segmentation cues (e.g. prosody, phonotactics etc.) in order to evaluate predictive power. If infants are able to do so, the relative accuracy of FWTPs and BWTPs can serve as a cue to morphological type and basic word order.

Local minima

Similarly to arbitrary global minima above, the results reported here refer to the Hungarian and the small Italian corpora. For the former, all four measures were evaluated against both WBs and MBs, and the combined FWTPs+BWTPs were also used to segment WBs and WIMBs separately.

Hungarian

Quantitative analysis The results are presented in Figure 3.26. BWTPs and MI turn out to be the best predictors on both the WB and MB criteria, with relatively high accuracy (85%) and completeness (50%–55%) scores. FWTPs are somewhat lower on both scores (50% accuracy and 75% completeness). FWTPs + BWTPs show radically different behaviors when compared against WBs and MBs. The former evaluation yields the most accurate segmentation (above 90%), however, with the lowest completeness (28%), whereas the latter evaluation produces similar and relatively high (around 80%) accuracy and completeness scores.

The results obtained by using a relative threshold correspond to the scores achieved by the 55^{th}–65^{th} percentiles of the absolute thresholds in terms of accuracy and the 45^{th}–55^{th} percentiles in terms of completeness.

Qualitative analysis All four measures segmented out numerous occurrences of most inflectional and derivational categories, just like in the case of global minima. What is most interesting about the local minimum algorithm is whether it allows to correctly segment ambiguous sequences. It has been found that frequent syllables such as *-ni* or *-ka, -ke* are correctly segmented when they appear as inflectional suffixes (infinitive marker and diminutive, respectively), and are correctly left unsegmented, at least in most cases, when they form part of a stem. This is not the case when both syllables in the ambiguous pair are highly frequent or highly infrequent. In the former case, a boundary is almost always posited (as in *le-nem*, which is typically segmented as a sequence of two functors ‘down’ and ‘no(t)’, even when it actually constitutes the last two syllables of the word *egyetlenem* ‘my one and only, i.e. my sweetheart/darling’). When both syllables are infrequent, no boundary is posited.

As expected, the relative thresholding mechanism does indeed provide more accurate segmentation in certain cases of ambiguity than absolute thresholding. However, the correct segmentation is not always found. Even if the ambiguity is accurately resolved for a syllable pair in most contexts, the wrong segmentation is proposed in a few others. It will be an important question for further research to establish what contexts are favorable

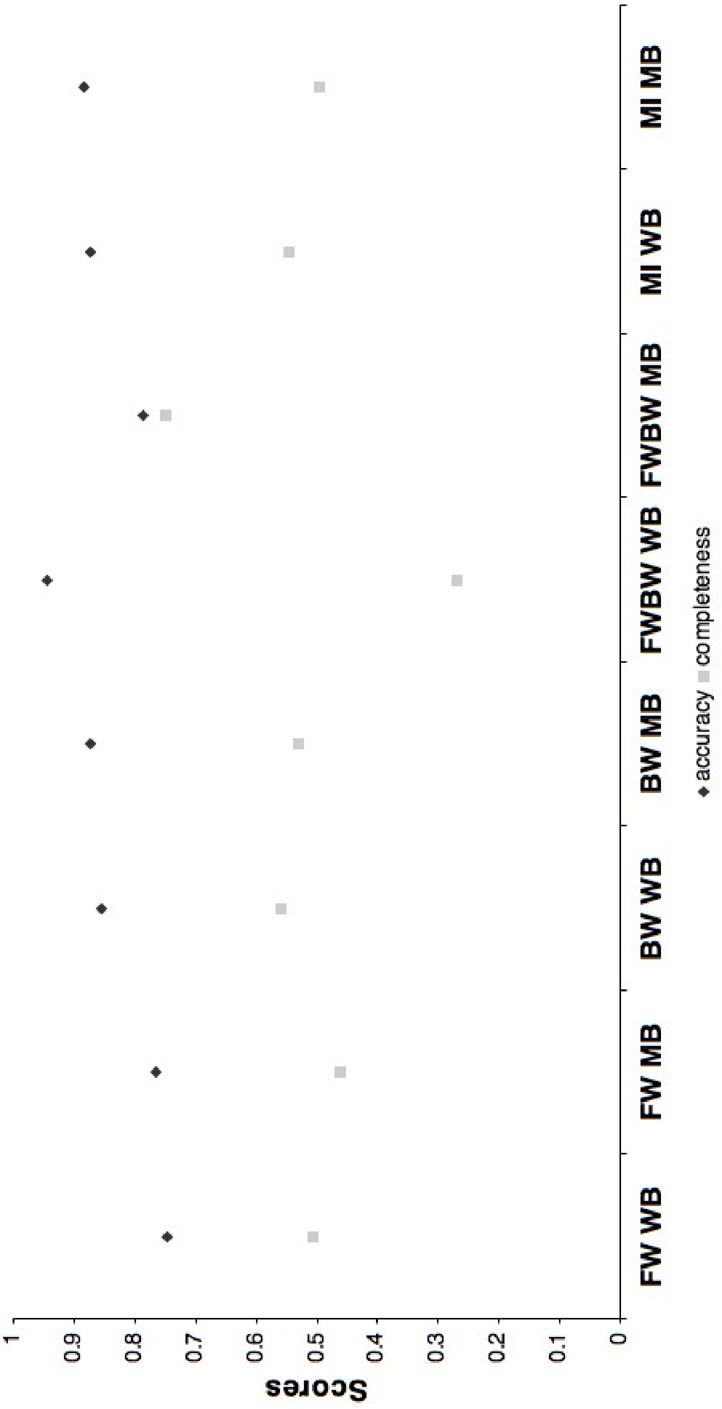


Figure 3.26: The accuracy and completeness of the segmentation obtained by using a relative threshold for the Hungarian corpus. The x-axis indicates the statistical measures and evaluation criteria. The y-axis shows the scores.

to accurate segmentation and why. The accuracy and flexibility of relative thresholding could most probably be enhanced if more contextual information were taken into account (e.g. two neighbors on both sides instead of one etc.).

Italian (Small corpus)

Quantitative analysis The scores are illustrated in Figure 3.27. FWTPs and MI perform best with this segmentation algorithm, achieving around 80% accuracy and 60% completeness. BWTPs score low both on accuracy (60%) and completeness (45%). FWTPs+BWTPs achieve very high accuracy (88%), but low completeness scores (30%).

Comparing them with absolute thresholds, these results correspond to the performance of the 50^{th}–70^{th} percentiles on accuracy and the 45^{th}–70^{th} percentiles on completeness.

Qualitative analysis Since the Italian corpus was not tagged at the morphemic level, much fewer cases of ambiguous segmentation were available. Those few that could be identified, e.g. *va-le* one word: 'is worth, valid'; or two words: 'go' and 'the.FEM.PL', were most often correctly segmented by relative thresholding.

Cross-linguistic comparison The results are replotted in Figure 3.28 for an ease of exposition. Similarly to the results obtained for absolute thresholds, FWTPs are better predictors for Italian than for Hungarian, while BWTPs favor the latter language over the other. FWTPs+BWTPs show a similar high accuracy, low completeness pattern in both languages, while MI patterns with the better predictor, i.e. with FWTPs in Italian and BWTPs in Hungarian.

Discussion Using a relative thresholding algorithm to segment the corpora, good results have been obtained. Except for the joint FWTPs+BWTPs, the results represent a kind of 'optimum' in the trade-off between accuracy

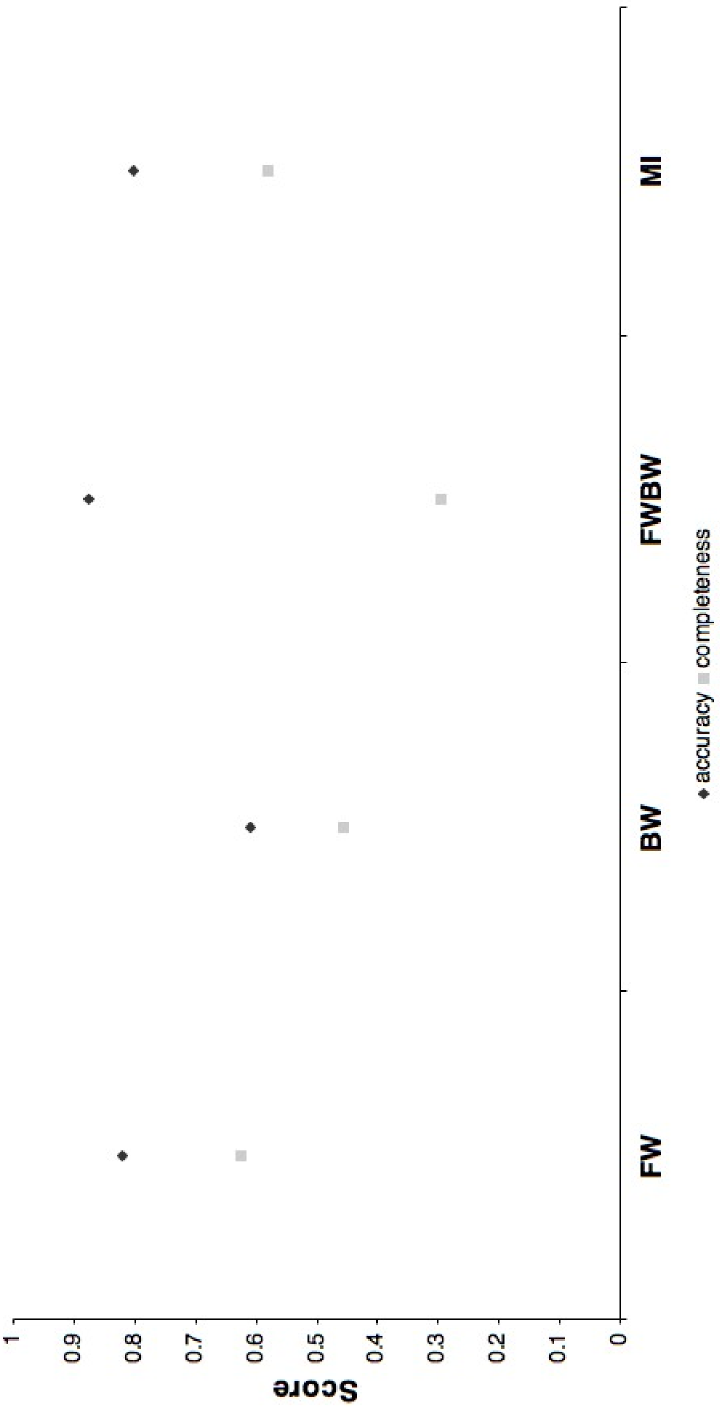


Figure 3.27: The accuracy and completeness of the segmentation obtained by using a relative threshold for the small Italian corpus. Graphical and plotting conventions are the same as in Figure 3.26.

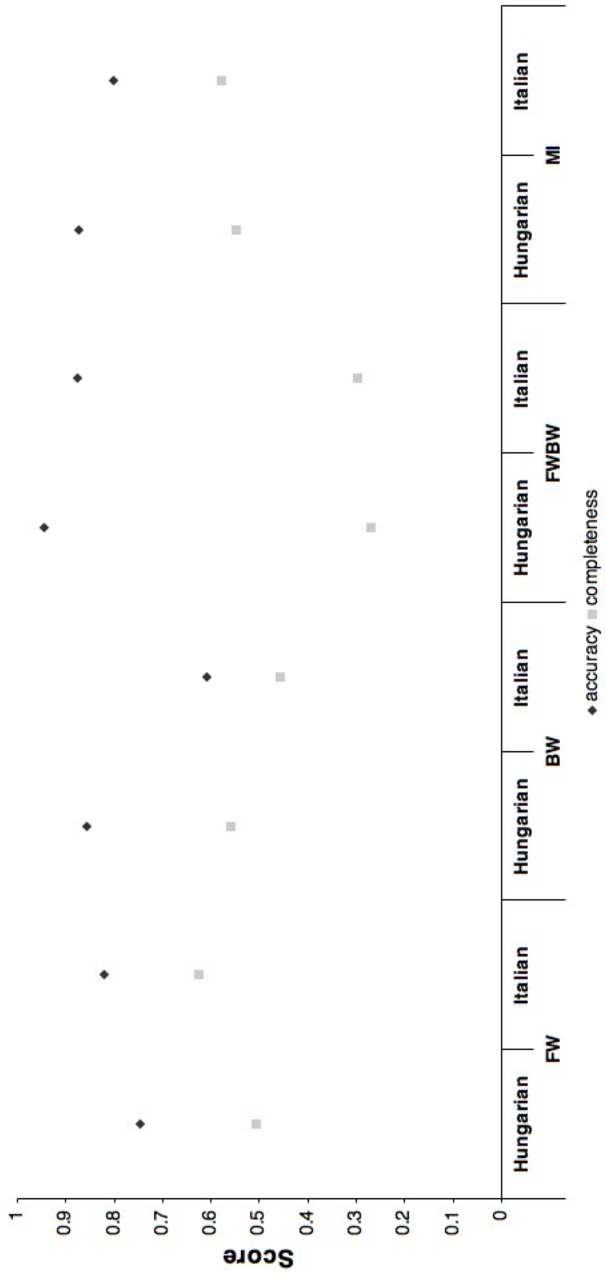


Figure 3.28: The accuracy and completeness scores obtained by using a relative threshold for the two corpora. Graphical and plotting conventions as in Figure 3.26.

and completeness. The number of boundaries predicted by the segmentation algorithm is high enough to produce a reasonable completeness score, but not too high to give rise to a large number of false alarms, thus accuracy is not compromised. This balance between accuracy and completeness is also reflected by the fact that the scores correspond to the results obtained for the 45th–70th percentile ranges of the absolute thresholds.

The exception to this general pattern is the combined FWTP+BWTP measure, which results in very high accuracy, but low completeness scores. The reason for this is that FWTPs+BWTPs, as discussed earlier, tend to be conservative (at least, when compared with WBs), since not only one, but two conditional probability values have to fall below the threshold. In the case of relative thresholding, this is further constrained by the fact that the values are compared to two other values instead of one predefined one. For FWTPs+BWTPs, relative thresholding thus involves two times two comparisons, which is a stricter criterion than in the case of the other measures.

The most important result obtained with the relative threshold algorithm is the difference between the predictive power of FWTPs and BWTPs in Hungarian and Italian. As with absolute thresholds, FWTPs have been found to be more reliable predictors of word boundaries in Italian, while BWTPs give better results in Hungarian. This, as discussed above, is a reflection of the agglutinating morphology and Complement-Head order of Hungarian, as opposed to Italian, which is inflecting and Head-Complement.

3.2.6 Discussion

In the current experiment, the segmentation performance of four statistical measures, FWTPs, BWTPs, FWTPs+BWTPs and MI, were tested using absolute and relative thresholding algorithms in two typologically different languages. Reasonably good segmentation was obtained in both languages with all four measures, but language-specific differences also obtained. Hungarian is better segmented using BWTPs, whereas Italian benefits from FWTPs. This difference is related to the morphological and syntactic characteristics of the two languages: Hungarian is rich in suffixes and phrase-final functors, while Italian has more frequent elements phrase-initially, e.g. prefixes etc. This finding converges with the results of Gammon (1969), who found that successor counts perform better than predecessor counts in English, which is a (weakly) inflecting VO language.

Importantly, it is not only the morphology of the languages that plays a role, but also its word order. This is not an unexpected result, since even intuitively, it is very probable that, say, a preposition will be followed by a determiner in English or in Italian. Note that these strong predictability relations have been found to hold within syntactic phrases, e.g. between preposition/postposition and determiner. It is an open question, worthy of further investigation, whether clausal constituent order also influences statistical distributions. Hungarian and Italian differ in this respect. In Italian, the basic word order is Subject-Verb-Object, with optional postverbal Subjects and *pro*-drop. Hungarian, in contrast, is a discourse-configurational language, where clausal constituents are ordered according to their pragmatic function instead of the syntactic one. The canonical order is Topic-Focus-Verb-Other. It will be interesting to test whether and if yes, how this difference is reflected in the statistics of the two languages.

While the above segmentation results are relatively good, a note of caution is in order. Since the average length of words is about 1.8 syllables, more than half of the syllable transitions are word boundaries. With an overgeneration strategy, i.e. by assuming that all transitions are boundaries, a completeness score of 100% and an accuracy of 51% and 49% can be achieved in Hungarian and Italian, respectively.[11] Obviously, this strategy does not require any statistics to be computed, so its results can be considered a 'lower boundary' or a minimum of what a segmentation has to achieve in order to be judged successful. When compared to this 'lower boundary', certain segmentations, e.g. BWTP in Italian, fare rather poorly. This minimum value also explains why accuracy plateaus around 50%-55% in the highest percentile ranges.

[11] This can be calculated from the standard equations for accuracy and completeness, if hits are taken to be the number of word boundaries, misses are 0 and false alarms are the number of word internal transitions.

In addition to segmentation, the overall distributions of the statistical measures were also computed and compared across-languages. No major difference has been found between them, except for those caused by the different sparsities of the corpora. Thus, it can be concluded that the distributions themselves cannot serve as a cue to morphological type.

In sum, to answer our initial developmental questions, conditional probabilities are good predictors of word and morpheme boundaries, at least under certain conditions, e.g. if a relative threshold or low absolute thresholds are used. In the higher value ranges, asymmetric conditional probabilities overgenerate boundaries. If only one type of boundary is posited, segmentation estimates word and morpheme boundaries equally well. The reason for this may be that only 11% of all boundaries were word internal morpheme boundaries in the Hungarian corpus. Thus, their segmentation does not modify accuracy and completeness scores considerably. When word boundaries and word internal morpheme boundaries are predicted separately by the algorithm, word boundaries are detected with much higher accuracy than word internal morpheme boundaries. This is because the latter are greatly overestimated.

3.3 Experiment 5: The role of frequency in cuing functors and content words in Japanese, Hungarian and Italian infant-directed speech

Besides conditional probabilities, frequency is also a commonly used statistical measure in computational and psychological studies of language. Therefore, its role in language acquisition needs to be assessed. A promising place to look for such contribution is the distinction between functors or grammatical words, such as determiners (*a, the, some* etc.), pronouns (*it, his, us* etc.) or prepositions (*of, on, over* etc.), and content words, such as nouns (*dog, boy, peace* etc.), verbs (*run, kiss, think* etc.), adjectives (*beautiful, good,* *new* etc.) or adverbs (*well, quietly, fast* etc.). These two superordinate categories of lexical items appear in all languages of the world, and have a fundamental functional role in the design of human languages (for a more detailed discussion, see Chapter 4). Among other surface features, functors and content words have been suggested to differ universally in their frequency distributions, functors having much higher token frequencies, but lower type frequencies than content words. This was indeed confirmed for English in several corpus studies. For instance, Cutler and Carter (1987) and Cutler (1993) report that functors made up 59% of the word tokens of their corpus, while they constitute only about 1% of all the word types, i.e. the lexical entries of English. Moreover, it has been observed that in English, there is little overlap in the frequency distributions of functors and content items. In Kucera and Francis's classical study (1967), the 50 most frequent lexical items were found to be function words.

If this observation carries over to other languages, then frequency can provide a universally reliable cue to category membership. Therefore, in this experiment, I test whether frequency reliably distinguishes functors and content words in Japanese and Italian infant-directed speech. Only Japanese and Italian were selected for this and the next Experiment, because they instantiate relevant typological options to compare.

3.3.1 Corpora

The Italian corpus was the same as the full size Italian corpus used in Experiment 4.

The Japanese corpus was derived from the infant-directed adult utterances of the Japanese Mother-Child Conversation Corpus collected at the Laboratory of Language Development, Brain Science Institute, RIKEN (Mazuka, Igarashi, & Nishikawa, 2006). I extracted all 22 mothers' utterances addressed to their infants (aged 18-24 months; 13 boys and 9 girls) in a laboratory environment during free play or directed story-telling (using specific story books) from the Mother-Child Conversation Corpus, excluding mothers' conversations with adults, e.g. the experimenter. The corpus thus comprises 14 958 utterances, made up of 47 071 word tokens, falling into 5205 word types[12] (see Table 3.1).

The corpus was purged of untranscribed material, but onomatopoeic words, sound imitations, fragments and other linguistic "noise" were kept unchanged under the assumption that they form a natural part of the input to young learners. All punctuation marks and spaces were deleted, except for utterance boundaries, which infants are known to be sensitive to (Jusczyk et al., 1992; Jusczyk & Kemler Nelson, 1996; Jusczyk, 1999) and thus can make use of during segmentation. The original corpus was coded using the Katakana syllabary with additional tags indicating details of actual pronunciation (e.g. vowel devoicing, palatalization of consonants etc.). The utterances in the infant-directed corpus used here were phonologically transcribed on the basis of the original enriched Katakana encoding in an automatic manner. The resulting phonological transcription was then checked by a native Japanese linguist.

Furthermore, the corpus had to be broken down into some representational units that statistics could be computed over. Existing results are not unequivocal about what form of representation infants use to represent and segment speech. However, there is some experimental evidence that goes in favor of the syllable. For instance, Mehler, Dupoux, and Segui (1990) showed that infants most readily represent speech as a sequence of syllables, and Saffran, Aslin, and Newport's (1996) results indicate that at least under experimental conditions, 8-month-olds are able to compute transition probabilities over syllables. Moreover, recent findings by Bonatti et al. (2005) suggest that adults cannot compute TPs over individual vowels (though they can over consonants). An additional practical advantage of choosing the syllable is that most previous studies (Gambell & Yang, 2004; Swingley, 2005) also used this unit, which makes the results more easily comparable. Consequently, the phonologically transcribed corpus was syllabified. Syllabification in Japanese is fairly straightforward and follows readily automatizable rules (see Section 3.1.2 above). The results were checked by a native Japanese

[12] In accordance with Japanese orthographic tradition, grammatical particles and markers were written and encoded as independent words.

linguist.

The syllabified corpus comprised 79 030 syllable tokens, falling into 1343 syllable types[13] (see Table 3.1). The average length of words measured in syllables was also calculated to check whether the Japanese corpus was subject to the monosyllabicity problem raised by Gambell and Yang (2004) in connection with infant-directed English. Words were found to be 1.68 syllables long, on average. They are, therefore, predominantly not monosyllabic.

3.3.2 Statistical measures

I calculated the frequency ranks (Zipf, 1935) of word types, and counted the number of functors and content words among the 100 most frequent words. I also computed 'overall coverage', i.e. what percentage of the corpora is covered by the most frequent functors and content words.

3.3.3 Results

Figure 3.29 illustrates the frequency distributions of the 100 most frequent words in the two languages. In Japanese, this list contained 57 functors and 43 content words. In Italian, the list had 67 functors and 33 content words. Importantly, as expected, in the highest frequency range, the distributions of the two categories are non-overlapping. Indeed, the most frequent Japanese content word is 21st in the rank (*Hora?* 'See?'), followed by only four other content words in the first third of the distribution (*So.* 'I see', *I!* 'Good! Great!', *mi* 'to see', *ko* 'child'). The most frequent Italian content word (*Guarda!* 'Look!') is 14th in rank, and it is followed by only two more content words in the first third of the distribution (*Fa!* 'Do!' and *mamma* 'Mum'). Note, in addition, that these content words are not used in a genuinely referential way. Rather, they function as phatic or discursive elements, e.g. forms of address (*mamma, ko*) and interjections (*Guarda!, So.* etc.).

I also calculated the cumulative token frequencies or overall coverage for the first 100 most frequent words, i.e. how much of the input they account

[13] Phonemically identical syllables were encoded as different if they carried different pitch accents.

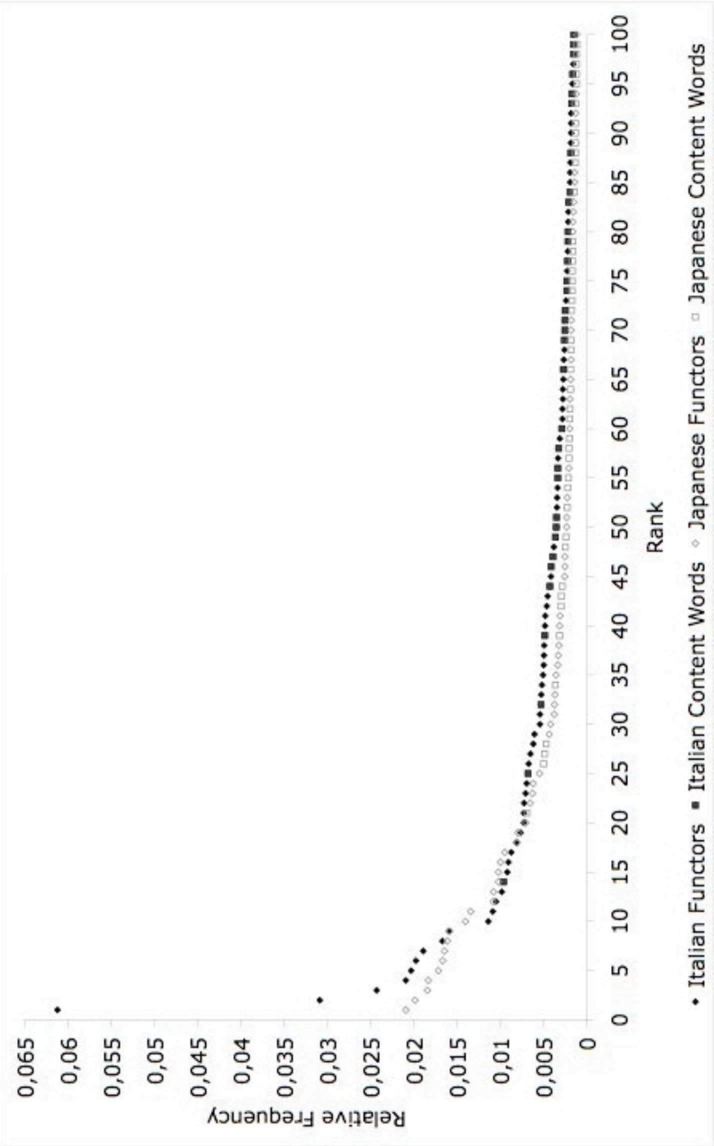


Figure 3.29: Histogram of the frequency distributions of the 100 most frequent words in the Japanese and Italian infant-directed corpora in Experiment 5. Light grey, empty markers represent Japanese functors (diamonds) and content words (squares). Black, filled markers represent Italian functors (diamonds) and content words (squares). The x-axis corresponds to the rank of a word in the frequency list. The y-axis shows relative frequencies (=absolute frequency/number of word tokens). Absolute frequencies could not be used, because the two corpora are not of the same size.

for. In Japanese, the 100 most frequent words altogether make up 47.45%, i.e. almost half of the corpus. Of this, functors make up 79.73% (corresponding to 37.83% of the whole corpus), content words 20.27% (corresponding to 9.62% of the whole corpus). Functors in the highest frequency range, i.e. the first third, where the distribution of the two categories is almost completely non-overlapping, account for 31.69% of the corpus. In Italian, the 100 most frequent words together account for 61.27% of all word tokens, i.e. almost two thirds of the whole corpus. Of this, functors make up 83.13% (corresponding to 50.93% of the whole corpus), content words cover 16.87% (corresponding to 10.33% of the whole corpus). Functors in the highest frequency range account for 39.26% of the corpus.

3.3.4 Discussion

From the above, it is clear that the frequency distributions of functors and content words show different patterns. As expected, individual functors occur more frequently than individual content words. Indeed, the thirty most frequent words are almost exclusively functors both in Italian and in Japanese, and these few functors account for about one third of the entire input that infants are exposed to. Therefore, frequency is a useful heuristic predictor of category membership.

3.4 Experiment 6: The role of frequency cuing word order in Japanese, Hungarian and Italian infant-directed speech

Functors and content words play clearly distinct roles in the design of grammar. Functors constitute the 'skeleton' of sentence structure, whereas content words add referents and lexical meaning. The distinction between the two categories might, therefore, provide learners with some initial insight into how grammar works. This is what I explore in this experiment.

Indeed, it has long been noted (Morgan, Shi, & Allopenna, 1996; Redington, Chater, & Finch, 1998; Mintz et al., 2002; Mintz, 2003) that functors tend to appear at salient sentential positions, i.e. at the edges of syntactic units. However, languages systematically differ in whether functors come at the left or the right edges of phrases. For example, Japanese, Basque or Turkish have postpositions, whereas English, Italian or French use prepositions. Importantly, it has been extensively documented in language typology (Greenberg, 1963; Dryer, 1992; Mehler et al., 2004) that the relative order of functors and content words correlates with a series of other word order phenomena, such as the basic word order of verbs and their objects, the order of complemetizers and subordinate clauses, or prepositions vs. postpositions, as illustrated before. It is precisely this empirical observation that is formally captured by the word order parameters of generative grammar.

Nevertheless, since young learners do not know where the boundaries of syntactic units lie within utterances—this is precisely what needs to be learnt—, this general information cannot be used to bootstrap structure. However, there is a special type of syntactic boundary that is available even to infants, namely utterance boundaries (Jusczyk et al., 1992). Therefore, I looked at the occurrences of functors and content words at utterance boundaries. Since Japanese is an OV language, frequent words are expected to appear phrase-, and thus utterance-finally, whereas in Italian, which is a VO language, frequent words were assumed to occur phrase-, and thus utterance-initially. If Japanese and Italian indeed exhibit opposite patterns, then it can be concluded that the order of frequent and infrequent words at utterance boundaries is a useful cue to bootstrap basic word order.

In the light of Experiment 5, functors were operationally defined as frequent words, content words were defined as infrequent words.

3.4.1 Corpora

The same Japanese and Italian corpora as in Experiments 4 & 5 served as the basis for this experiment. However, since one-word utterances are not informative about word order, I discarded these, and extracted only multiword utterances. With this manipulation, I obtained a corpus of 9889 utterances in Japanese and 42 955 utterances in Italian.

3.4.2 Statistical measures

I used the multiword utterances of the corpora to calculate how often frequent and infrequent words appear at initial and final positions at utterance boundaries. Frequent and infrequent words (FW and IW) were defined as having a (relative) frequency of occurrence higher and lower, respectively, than 4 predefined thresholds: T1=0.01, T2=0.005, T3=0.0025 and T4=0.001. T1 defines 12 words as frequent in Italian, and 20 in Japanese, roughly corresponding to the highest frequency range where only functors appear. T2 defines 36 words as frequent in Italian, and 34 in Japanese, still corresponding to the frequency ranges where there is little overlap between the distributions of the two categories. T3 defines 72 words as frequent in Italian, and 63 in Japanese. Finally, T4 defines 133 words as frequent in Italian, and 144 in Japanese. All other words in the corpora were categorized as infrequent. For further descriptive statistics about the four thresholds, see Table 3.2. No lower thresholds were used, because the words categorized as frequent by T4 already covered about two thirds of the corpora. Decreasing the threshold further would have rendered the frequent/infrequent distinction meaningless, as almost all words would have been categorized as frequent.

Using the frequent and infrequent categories as defined by the four thresholds, I calculated the percentages of the different possible word orders at the boundaries of multiword utterances. These measures were obtained in the following way. The first two and the last two words of all utterances, that is two-word 'phrases' at the left and right utterance boundaries, were identified. If the 'phrase' had a [FW IW] order, it was counted as 'frequent-initial'. If it had an [IW FW] structure, it was counted as 'frequent-final'. 'Phrases' where both words were of the same category, i.e. [FW FW] or [IW IW] did not enter into the counts , as they were not informative about the relative order of frequent and infrequent words. Since the two corpora were not of equal size, the counts were transformed into percentages.

To evaluate the results statistically, we divided both corpora into 10 equal-

	Number of 'Frequent Words'		Frequency Threshold		Coverage	
relative frequency threshold	Japanese	Italian	Japanese	Italian	Japanese	Italian
T_1=0.01	20	12	471	2457	29%	26%
T_2=0.005	34	36	236	1173	38%	43%
T_3=0.0025	63	72	119	595	48%	56%
T_4=0.0001	144	133	48	235	61%	65%

Table 3.2: Some quantitative properties of the category 'frequent word' in Japanese and Italian, as defined by the four different relative frequency thresholds used in Experiment 6. The 'Number of 'Frequent Words" gives the number of word types in the frequent word category. The 'Frequency Threshold' gives the value of the four thresholds in terms of absolute frequencies. 'Coverage' indicates what percentages of the corpora are accounted for by 'frequent words'.

sized subcorpora, calculated the percentages for the individual subcorpora using all 4 thresholds, and conducted ANOVAs over these datasets. We expected to find an interaction between languages and word orders, as an indication of opposite word orders in Japanese and Italian.

3.4.3 Results

Figure 3.30 presents the percentages of frequent-initial and frequent-final utterances in the two languages using the 4 different thresholds. As expected, Japanese and Italian show the opposite patterns. Japanese has more frequent-final utterances, while Italian has more frequent-initial ones. Numerically, T1 identifies 47% of the multiword utterances as frequent-final and 27% as frequent-initial in Japanese, 25% as frequent-final and 54% as frequent-initial in Italian. T2 identifies 55% as frequent-final, and 31% as frequent initial in Japanese, and 26% as frequent-final and 64% as frequent-initial in Italian. T3 identifies 54% as frequent-final and 31% as frequent-initial in Japanese, 26% as frequent-final and 66% as frequent-initial in Italian. T4 identifies 46% as frequent-final and 29% as frequent-initial in Japanese, 24% as frequent-final and 62% as frequent-initial in Italian.

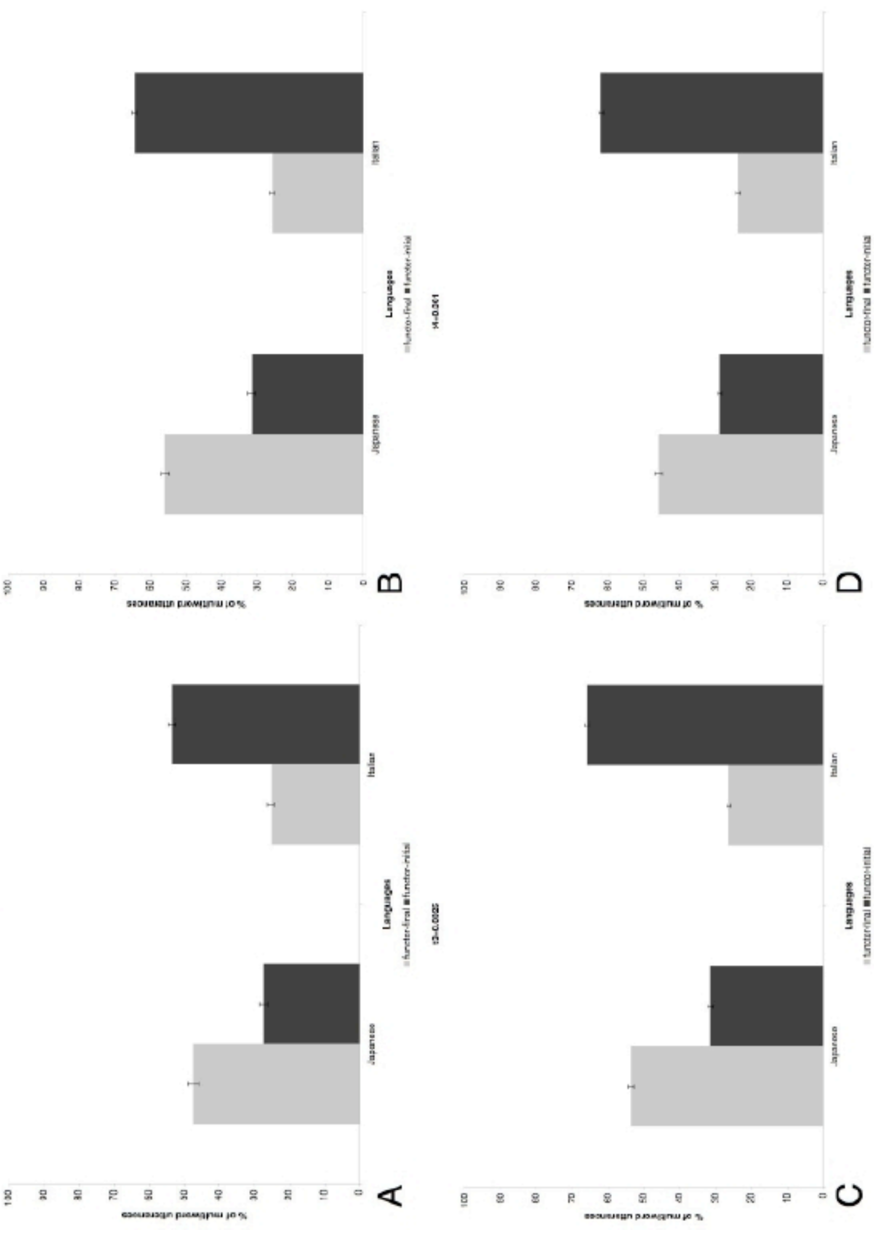


Figure 3.30: The percentage of frequent-initial and frequent-final phrases at utterance boundaries in the Japanese and Italian infant-directed corpora, using four different relative frequency thresholds to define frequent words. Panels A-D show the results at the four different thresholds. Light grey bars represent frequent-final phrases. Dark grey bars represent frequent-initial ones. The y-axis corresponds to the percentage of multiword utterances. Errors bars show standard errors of the means.

We carried out an ANOVA with factors Language (Japanese/Italian) and Order (frequent-initial/frequent-final) for each threshold using the percentages of frequent-initial and frequent-final 'phrases' in the 10 subcorpora as the dependent measure. For T1, we obtained no main effect of Language. But there was a significant main effect of Order ($F(1,39) = 37.822, p < 0.001$), indicating that there were more frequent-initial phrases in the two corpora together than frequent-final ones. Crucially, there was a significant interaction Language X Order ($F(1,39) = 1311.3, p < 0.0001$) due to the opposite order patterns attested in the two languages. For T2, the ANOVA showed no main effect of Language, but there was a significant main effect of Order ($F(1,39) = 59.560, p < 0.0001$), once again reflecting the fact that there were more frequent-initial phrases overall in the two languages than frequent-final ones. Just as before, we also obtained a significant interaction Language X Order ($F(1,39) = 1161.6, p < 0.0001$), indicating that Japanese had more frequent-final phrases, while Italian had more frequent-initial ones. Unlike in the previous two cases, the ANOVA for T3 revealed a significant main effect of Language ($F(1,39) = 42.118, p < 0.001$), indicating that this threshold filtered in more sentences in the Italian corpus than in the Japanese one. In addition, as before, we also found a significant main effect of Order ($F(1,39) = 135.60, p < 0.0001$), as well as a significant Language X Order interaction ($F(1,39) = 1709.0, p < 0.00001$), indicating opposite orders in the two language. Using T4, a similar pattern was obtained, with a significant main effect of Language ($F(1,39) = 72.158, p < 0.0001$) and Order ($F(1,39) = 178.74, p < 0.0001$), and a significant interaction between the two factors ($F(1,40) = 1213.3, p < 0.0001$) .

3.4.4 Discussion

These results confirm the prediction that the relative order of frequent and infrequent words is the opposite in Italian and Japanese, as expected on the basis of the theoretical characterization of word order in these languages. Italian has more frequent-initial phrases at utterance boundaries than frequent-final ones, while Japanese has more of the latter type. This observation is true in both languages irrespectively of how FWs were defined, i.e. what frequency threshold was used. In addition to this relative difference between word orders, the absolute numbers of word order types are also informative. Indeed, in Italian, utterances starting or finishing with a frequent-initial phrase constitute the absolute majority of all utterances at any of the four thresholds. In Japanese, utterances with frequent-final phrases outnumber other utterances at all four thresholds, and reach absolute majority at two of them. Thus, as expected, the relative order of frequent and infrequent items at utterance boundaries is a strong predictor of the basic word order pattern of a language.

Interestingly, in addition to the above finding, two more results were obtained: (i) overall, there were more frequent-initial utterances than frequent-final ones (at all thresholds), and (ii) at thresholds T_3 and T_4, more sentences were identified for Italian than for Japanese. Our assumption is that these results are at least partly attributable to another word order property, namely the relative order of determiners and nouns (formally, the Specifier-Head parameter). The order is [Det(erminer) N(oun)] in both languages (Japanese: *kono hon* 'his book'; Italian: *la tavola* 'the table'). However, Japanese has fewer determiners than Italian. Importantly, it lacks articles altogether, and only has demonstratives, numeral classifiers etc. Since determiners are functors, thus frequent words, while nouns are less frequent content words, the [Det N] pattern, common to both languages, increases the overall amount of frequent-initial utterances. Additionally, given that in Italian, there are more determiners than in Japanese, the lower thresholds identified more utterances in the former language than in the latter. This does not happen for the two higher thresholds, because the most frequent functors are not determiners, and thus follow the Head-Complement, rather than the Specifier-Head pattern. In addition, other factors, such as the more varied nature of the Italian corpus, might also contribute to the presence of the additional effects. Whatever the definitive explanation may be, these effects are much smaller than the effect of the opposite word orders (cf. the statistical results above), thus they do not blur the strong interaction between language and word order that we are focusing on here.

In sum, the data shows that frequency information in the input can be used as a heuristic predictor of category membership. This information, in turn, can be used to extract basic word order from phrases at utterance boundaries.

3.5 General Discussion: Statistical information in the input to infants

The three experiments presented in this chapter have investigated some aspects of the statistical information contained in the signal young learners receive. In particular, infant directed corpora in three unrelated and typologically different languages, Japanese, Hungarian and Italian were studied. The main conclusion of the investigations is that statistical information contained in the signal provides cues to several morphosyntactic properties of languages, such as agglutinating type, lexical categories and basic word order.

Frequent items appear to play a critical role in all of the above. Conditional probabilities segment words out in the order of their frequencies, so frequent items get identified first. The position of frequent items with respect to utterance boundaries indicates basic word order. As it has been shown, frequent items correspond to the functors of a language. It seems, then, that functors may contribute to the acquisition of morphosyntax more importantly than it has been assumed so far.

While the experiments have shown that statistical information can signal many different linguistic properties, this does not imply that statistical learning alone is sufficient to explain the acquisition of morphology and syntax. Without a syntactic representation of word order, e.g. the Head-Complement parameter or some equivalent, the relative order of frequent and infrequent items at utterance boundaries remains just that—the relative order of frequent and infrequent items at utterance boundaries. For this information to allow infants to learn something about, say, the order of prepositions and nouns, the statistical cues need to be mapped onto a linguistic representation.

Even if the signal is rich in statistical information, it has to be shown that infants are sensitive to it and use it during language acquisition. For conditional probabilities, (Saffran, Aslin, & Newport, 1996) and subsequent work has provided ample evidence that babies can use TPs to segment continuous streams. However, the use of frequent functors as indicators of word order is not known. Chapter 4 will take up this question.

Chapter 4

What is in learning? Bootstrapping lexical categories and word order: a cross-linguistic artificial grammar learning study in adults and prelexical infants

Imagine for a moment that you are an English-learning 8-month-old, hearing the following speech sequence:

(1) . . .alicewasbeginningtogetverytiredofsittingbyhersisteronthebank. . .[1]

In order to make some sense of this, you will have to break the continuous stream up into its constituent units, words etc. However, as a young learner, you don't yet know the words of English, not to mention its syntax and morphology. So what can you do? There are a few chunks that you have heard on and on again (Chapter 3, Experiment 5), for example, /ðə/, /əv/, /iz/ etc., which you cannot help recognizing.

[1] The opening sentence of Lewis Carroll: *Alice's Adventures in Wonderland.*

(2) . . .alice**was**beginning**to**getverytired**of**sitting**byher**sister**onthe**bank. . .

So what if you used these as your breakpoints? Why not chop the stream up where these 'words' appear? All you have to decide is whether you cut before or after these 'words'. Since you have very often heard them at the beginning of what people say, but not at the end, (Chapter 3, Experiment 6), a good bet seems to be to put your boundaries before them. So you would get something like this:

(3) . . .[alice] [**was**beginning] [**to**getverytired] [**of**sitting] [**byher**sister] [**onthe**bank] . . .

It turns out that you would not be too much off the mark. The pieces you would end up with are not very different from the actual underlying syntax of this English sentence. The Chapter looks at whether this strategy, which I will term the frequency-based bootstrapping of word order, has any plausibility in real language acquisition.

Chapter 3 has shown that the input offers at least some reliable cues, namely frequency and position with respect to utterance boundaries, to indicate lexical categorization and basic word order. The present Chapter investigates whether learners use these cues, by testing whether adults and infants, when faced with an artificial, thus unknown language, form an expectation about its word order on the basis of the patterns found in their native language. Crucially, if infants can be shown to have such representations very early on, even before they build their lexicon, it constitutes strong evidence that frequency patterns play a role in bootstrapping word order.

To lay the groundwork, I first discuss linguistic and neuropsychological evidence demonstrating that functors and content words are two distinct categories, and that this distinction plays a crucial role in the design of language. Then, I present a first series of artificial grammar learning experiments to show that adult speakers of five, typologically different languages, Basque, Japanese, Hungarian, Italian and French, do indeed use their different representations of word order to learn novel linguistic material. Finally, I report a similar experiment comparing the word order expectations of 8-month-old, i.e. prelexical, Japanese and Italian babies.

4.1 The distinction between function words and content words

Fundamental to language is the distinction between functors and content words. The former carry lexical and referential meaning, while the latter encode grammatical relations. This functional difference constitutes an essential and universal design feature of all human languages.

4.1.1 Defining the functor/content word distinction

It has been argued at least as long back as Dionysius Thrax (100 B.C.) that lexical items cluster into categories. The most basic among these is the distinction between function words and content words, based on the different roles they play in syntactic and semantic computations.

In current linguistic theory (Chomsky, 1995), lexical items are conceived of as bundles of phonological, syntactic and semantic features, i.e. properties. Under this view, functors and content words play different roles in linguistic computations because they are made up of syntactic and semantic features of different nature. The syntactic features of functors motivate syntactic operations (i.e. trigger movement, agreement etc.), whereas the syntactic features of content words do not (although they might enter into syntactic computations initiated by function words). As for the semantic properties, content words carry rich lexical meaning, while function words signal grammatical relations and do not refer to entities in the world.[2] As a natural consequence of this difference, content words constitute open, extendable classes, since new objects appear every day, while function words form closed classes.[3] Typically, the loss or the introduction of functors entails

[2] In more technical terms, content words typically denote (sets of) entities or (zero-order) relations between sets of entities. Function words, on the other hand, denote higher order functions, i.e. functions over functions. The semantic distinction is less clear than the syntactic one, though.

[3] Note, however, that the functor/content word distinction does not exactly coincide

a change in the historical/genealogical development of a language, whereas no such process is implied when content words are added or lost. Although languages differ with respect to which universally available content or function word subcategories they grammaticalize and how they implement them, the major divide between function and content words has been shown to be universal (Abney, 1987; Fukui, 1986).

Probably the most insightful illustration of the distinction is given by Lewis Carroll's *Jabberwocky* poem: *'Twas brillig, and the slithy toves \ Did gyre and gimble in the wabe . . .* The content words are replaced by novel tokens, while the grammatical structure is maintained by leaving function words in place.

It has long been observed that functors tend to be shorter and more reduced than content words, e.g. (Selkirk, 1984; Nespor & Vogel, 1986).[4] Morphologically, function words tend to be simple, i.e. typically monomorphemic, and usually cannot undergo derivation (e.g. *to* → **to-ing*, **to-ness*, **to-ity*). At the level of the word, functors tend to contain a minimal number of syllables, and very often lose their independent word status (e.g. *it is* → *it's*; for a formal characterization of this procedure, see (Selkirk, 1996)). The syllables that make up function words also appear to be minimal, with no or simple onsets, non-diphthonguized nuclei and no or simple codas. In addi-

with the open class/closed class one. Numerals or certain prepositions/postpositions constitute closed classes, yet are relatively contentful in meaning. Nevertheless, the two distinctions overlap to a great extent and are often used interchangeably in the literature—a practice that I will not depart from either.

[4] It has been suggested that there might actually be a connection between the phonological minimality and high token frequency of functors. Intuitively, the proposal is that the more frequent or probable an event is, the less information is needed to identify and recognize it. In computational linguistics, this intuition has recently been formulated as the Probabilistic Reduction Hypothesis: "word forms are reduced when they have a higher probability" (Jurafsky, Bell, Gregory, & Raymond, 2000). As preliminary evidence for the hypothesis, Jurafsky and colleagues have measured several phonological, phonetic and acoustic features (vowel reduction, word length, final consonant dropping, etc.) of function and content words in telephone conversations. They have shown that high frequency and high predictability, measured in terms of the relative frequency of the target word, the transitional probability between the target word and the previous and/or following word(s) and the joint probability of the target word and the previous and/or following word(s), were all good predictors of reduction for functors. Content words, however, were reduced only when they were highly frequent, but not when they are highly probable.

tion, their phonemes are often subject to reduction or underspecification (e.g. the vowels of English function words are frequently centralized to schwa). In tone languages, tones of function words are reduced or altered in predictable, context-dependent ways. Acoustically and phonetically, function words are characterized by short duration, low pitch and low amplitude.

Recently, Morgan and colleagues (Morgan et al., 1996) have shown that these minimality and reduction effects are also present in infant-directed speech. They conducted a systematic cross-linguistic study of the perceptual differences between function and content words in infant-directed English and Mandarin Chinese, measuring a series of phonological and acoustic measures (e.g. number of syllables, syllable complexity, diphthonguization, vowel duration, amplitude etc.). They have found that function words are perceptually *minimal* on all of these measures, while content words are not. Interestingly, Morgan and colleagues have found that no perceptual cue alone is sufficient for good categorization (they provide about 60% precision each). However taken together, they offer reliable indications (about 80-90% precision). This is because not all languages implement all the cues (e.g. English has no tones), and even among the implemented cues, not all are contrastive in every language (e.g. in Chinese, not only function words, but also content words are monosyllabic). In other words, no perceptual cue is universally valid in itself. However, some subset of them should provide enough information for correct categorization in any language.

There is a large body of psycho- and neurolinguistic evidence to support the distinction between functors and content words in the mature language faculty. The two categories have been found to play different roles in sentence processing (Pollack & Pickett, 1964; Garrett, 1975; Bradley, 1978; Segui, Mehler, Frauenfelder, & Morton, 1982; Gordon & Caramazza, 1982, 1985; Cutler & Carter, 1987; Cutler, 1993; Friederici, 1985; Herron & Bates, 1997). They show different ERP[5] signatures (Neville, Mills, & Lawson, 1992; Nobre, Allison, & McCarthy, 1994; Pulvermüller, Lutzenberger, & Birbaumer, 1995; Osterhout, Bersick, & McKinnon, 1997; King & Kutas, 1998; C. Brown, Hagoort, & Keurs, 1999), and they doubly dissociate in aphasic (Gardner

[5] For abbreviations, see p. 1.

& Zurif, 1975; Caramazza & Zurif, 1976; Friederici, 1985; Keurs, Brown, Hagoort, & Stegeman, 1999; Keurs, Brown, & Hagoort, 2002) and dyslexic (Caramazza, Miceli, Silveri, & Laudanna, 1985; Silverberg, Vigliocco, Insalaco, & Garrett, 1998; Druks & Froud, 2002) populations.

4.1.2 Functors and content words in infants' linguistic knowledge

Although it is a well known fact (Guasti, 2002) that young children very often omit functors in their early productions, several studies have addressed the issue of whether they are, nevertheless, able to pick up the low-level cues and represent functors. An early study (Shipley, Smith, & Gleitman, 1969) showed that children whose linguistic production was at the 'telegraphic' phase (it contained no function words), nevertheless understood instructions better if the instructions themselves were not telegraphic, but contained function words as well. Later Gerken and her colleagues (Gerken, Landau, & Remez, 1990) established that the omission of functors in early production stems from a limitation on production, and not on perception or encoding. In a series of imitation experiments with 2- to 3-year-old children, they found that children tend to omit weak, unstressed monosyllabic morphemes (the equivalents of functors), but not strong, stressed ones (content words), even if both are nonsense, non-English words. Also, they imitate non-existing content words with greater ease if they appear in the environment of English function words as opposed to environtments of nonsense function words. Moreover, children make a distinction between those nonsense functors that follow the usual consonant patterns of English function words and those that do not. Taken together, these results indicate that even though young children might have problems producing functors, they still build fairly detailed representations of them, which, then, they can use in *segmenting* and *labeling* the incoming speech stream. In a later experiment, Gerken and McIntosh (Gerken & McIntosh, 1993) obtained similar results for sentence comprehension. They presented short imperative sentences of the kind *Find* ***the*** *bird for me* to 2-year-old children in a picture selection task, where subjects were asked to choose the right picture (e.g. the bird) out of four alternatives. The article in front of the noun in the imperatives was manipulated in the following way. It could be (i) the grammatically correct English function word, i.e. the article *the*; or (ii) an ungrammatical, but existing English function word, *was*; or (iii) a phonologically function word-like, but non-existent morpheme, *gub*; or (iv) omitted altogether. Children performed significantly better when the grammatical function word rather than the ungrammatical or the nonsense one was used, indicating that at 2 years of age, children are already aware of the distributional properties of at least certain function words. Importantly, this was true even in the subgroup of subjects whose mean length of utterance was below 1.50 (that is, they basically produced one-word speech), provided that a high-pitched female voice was used to produce the stimuli.

However, the above experiments were carried out with children who have already broken into the structure of their native language. But segmentation and labeling cues are the most necessary at the beginning of acquisition to break up the input. Indeed, Shi, Werker, and Morgan (1999) asked whether newborns are able to distinguish the phonological cues correlated with the two categories. Their findings indicate that newborn infants of both English-speaking and non-English-speaking mothers are able to categorically discriminate between English function and content words when those were presented in isolation. At 6 months of age, infants start to show a preference for content words (Shi & Werker, 2001). By 11 months, though not yet at 8 months, they are also able to represent frequent functors, e.g. *the*, but not infrequent ones, e.g. *their*, in some phonological detail (Shi, Cutler, Werker, & Cruickshank, 2006; Shi, Werker, & Cutler, 2006). However, they are able to use functors, frequent and infrequent alike, at both ages to segment out a following content word (Shi, Cutler, et al., 2006). Similar findings have been obtained by Höhle and Weissenborn (2003), who observed the same categorical discrimination in 7- to 9-month-old German infants exposed to continuous speech.

Moreover, Shafer, Shucard, Shucard, and Gerken (1998) have shown that infants at 11 months of age, but not yet at 10 months, show different ERP signatures upon hearing unchanged continuous English speech versus a continuous English speech in which a tone was superimposed on the function words resulting in a substantial distortion of their acoustic/phonetic characteristics.

On the basis of the above, it is not unreasonable to assume that the functor/content word distinction is available to infants very early on. What role, if any, might this distinction play during acquisition?

4.1.3 Functors' role in the learnability of language

Lexical categorization

A deep-rooted intuition in the structuralist–generativist tradition is that functors are fundamental for the categorization of content words. Inspired by Chomsky's (1957) early ideas, Thorne (1968) developed a learning and parsing model that made use of some *a priori* syntactic knowledge and a limited lexicon. He found that for successful learning and parsing, the optimal composition of the lexicon was one that a priori contained all the function morphemes, both free and bound, but only a small number of content words.

Redington et al. (1998) arrived at similar results. They ran a series of nine experiments testing different aspects of categorization over a very large infant-directed English subcorpus of the CHILDES database. They obtained high accuracy (79%) and somewhat lower completeness (45%) results. The distributional context was defined as a window of two words before and after the target word. Both target words and context words were taken from the n most frequent words of the corpus. When the size of n was systematically varied for both target and content words, categorization yielded inverted U-shaped curves, with an optimum of around 1000 target words and 150 context words. The authors investigated whether all lexical categories benefit equally from distributional evidence, and found a clear asymmetry between function words and content words. Content words were categorized with fairly high accuracy (ranging from 38%–90%) and lower completeness (18%–53%), while function words showed the opposite pattern (accuracy range: 9%–33%; completeness range: 24%–100%). Therefore, distributional analysis requires a relatively small number of context elements, which cannot themselves be learnt through distributional analysis. These are typically functors. The essential role of function words for categorization was further shown in two other manipulations. In one condition, all function words were removed from the corpus, which, as expected, greatly impaired categorization. In another condition, all function words were replaced in the corpus by the label FUNCTION. This manipulation had a slight negative effect on performance.

Mintz et al. (2002) brought this idea further. After replicating the distributional analysis, they introduced two modifications to directly ask acquisition-related questions. First, since infants are known to detect phonological phrase boundaries (Christophe, Dupoux, Bertoncini, & Mehler, 1994; Jusczyk & Kemler Nelson, 1996), the authors redefined distributional contexts so as to reflect phrasal bracketing. This manipulation resulted in phrases introduced by a functor, followed by content words. With this bracketing in place, a better categorization was obtained for both nouns and verbs, the two lexical classes the study focused on. Implemented independently of the first modification, the second consisted in collapsing all functors into one category and replacing them with a uniform label, similarly to Redington et al. (1998). Contrary to the findings of this latter study, Mintz et al. (2002) have found no decrease for noun categorization, and an actual improvement for the classification of verbs.

The picture that emerges from the above computational studies is the following. Given the statistical properties of the input, lexical categorization is most efficient if it is bootstrapped through other mechanisms that establish the initial categories. The most readily available and at the same time the most informative candidates are frequent functors. They provide the background relative to which other categories, mainly categories of content words, can be established.

Rule learning

Another role that functors have been assumed to play in language acquisition is to cue rules and increase the learnability of structural generalizations. This hypothesis has been explored in a number of artificial grammar learning studies (Braine, 1966; Green, 1979; Morgan & Newport, 1981; Mori & Moeser, 1983; Morgan, Meier, & Newport, 1987; Valian & Coulson, 1988; Valian & Levitt, 1996), asking whether (artificial) languages with and without the functor/content word distinction show different degrees of learnability.

Braine (Braine, 1963, 1966) was one of the first to study how frequent or constant marker elements influence grammar learning. Sequential, linear order is a fundamental aspect of natural languages. However, very often what is important in a grammatical construction is not the absolute position in a sequence, but the position of an element with respect to another one. Consequently, it is important to know whether humans are able to learn languages on the basis of information about relative rather than absolute position. Braine (1966) tested this in 9-10-year-old children, giving them artificial grammar learning tasks in which success depended on learning the positions of non-frequent variable tokens ('content words') with respect to constant marker elements ('function words'). The positions to be learnt could be immediately adjacent to or one position removed from the marker element (P and Q, respectively, in fPQ, where f is a marker). The results suggest that subjects readily learn both relative positions. This, as the author points out, is a necessary prerequisite for natural language acquisition.

Green (Green, 1979) investigated the importance of the reliability of functors as markers. In a first experiment, he visually presented three different grammars to three groups of subjects. The first group saw well-formed strings (ones that obeyed certain sequential ordering rules) from a grammar containing functional markers and content words, which co-occurred in a systematic way ('effective markers'). The second group was familiarized with a grammar having markers and content words, but they co-occurred randomly ('useless markers'). The third group was presented with a grammar having only content words and no markers at all ('no markers'). The author found that there was some learning in all three conditions, but learners of 'effectively marked' grammars performed significantly better than subjects in the other two conditions. In a second experiment, similar grammars were used, but in addition to the word category markers, phrasal category markers were also introduced. A phrase was defined as a sequence consisting of a word marker, a content word from the corresponding category, a second word marker and a second content word from the corresponding category, e.g. [a A b B]. Phrases, just like words, came in categories, and each phrasal category had its respective phrase marker, e.g. [p a A b B]. Six different grammars were tested: (i) one with all the markers in place, (ii) one with no phrase marker distinctions, i.e. the same phrase marker for all phrasal categories, (iii) one with no between-phase word marker distinctions, (iv) one with no within-phrase word marker distinctions, (v) one with no word marker distinctions at all, i.e. the same word marker in all positions irrespectively of word category, and (vi) one with no effective markers at all, i.e. the same uniform marker in all word and phrase marker positions. Subjects were tested on sentence completion tasks with the final word or the final phrase missing, plus on a serial ordering task, in which they had to rearrange jumbled words in the grammatical order. The results showed that subjects perform better when all the markers were in place, and showed selective problems on the word or on the phrase completion task depending on which kind of marker was undistinguished in their grammars. Green (Green, 1979) synthesized these findings in the 'marker hypothesis', which has the following three tenets. First, in all learnable languages, there will be a small set of words or morphemes, the 'markers', each of which is associated with one or, at most, a few syntactic constructions/categories. Second, sentences are easier to parse, when they contain markers. Third, a language without markers would be very difficult or impossible to understand. One consequence of the marker hypothesis is that a language that is hard or impossible to parse is also hard or impossible to learn. Thus, Green posits an indirect relation between the existence of markers and learnability, mediated by parsing constraints.

Morgan and colleagues (Morgan et al., 1987) conducted similar experiments, comparing learning in artificial grammars which had (i) no markers, (ii) inconsistent markers or (iii) consistent markers. They focused mainly on how, if at all, markers help learners discover the hierarchical phrase structure of the input. Importantly, they tested free and bound functors, i.e. function words and grammatical suffixes, separately. In the experiment that tested free function words, three grammars were used. One ('no markers') contained only content words, and no function words in the first condition. A second ('inconsistent markers') used both function words and content words, but in such a way that function words did not mark phrase boundaries, rather they appeared randomly between content words. A third ('consistent markers') had both function words and content words, in such a way that function words indicated real phrase boundaries. Apart from the functors, all three grammars were generated by the same phrase structure rules. Subjects were tested on two tasks. The first checked for the learning of the linear order and sequential co-occurrence patterns of content word categories. The second examined the correct induction of the hierarchical constituent structure. The results showed that subjects learned the linear order and sequential co-occurrence patterns in all conditions. However, those in the consistent markers condition performed better than the others. Moreover, only they succeeded in the constituency tests. The second experiment, using bound function morphemes, was very similar to the previous one. In the first condition ('no markers'), no grammatical suffixes were added to the content words; the sentences were actually identical to those in the no marker condition of the previous experiment. In the second condition ('inconsistent markers'), the same suffix was added randomly to content words that did not belong to the same phrase. In the third condition ('consistent markers'), the suffixes were attached to content words that constituted the same phrase. The results were similar to the ones obtained before. All subjects performed well in the order and co-occurrence tests, with the consistent markers group outperforming the other two, while only the consistent group succeeded in the constituency test, although the difference between them and the two other groups was much smaller here than in the previous experiment. Morgan and colleagues conclude that markers, both free and bound, provide efficient cues to hierarchical phrase structure.

Austin, Newport, and Wonnacott (n.d.) have recently reported very interesting findings suggesting that inconsistency in the input is processed differently by young children and adults. The authors exposed adults and children to one of two artificial grammars. Both used two determiners, one occurred 2/3 of the times, the other 1/3 of the times. However, in the 'consistent' grammar, the two determiners were lexically conditioned, each preceding its respective set of nouns, whereas in the 'inconsistent' grammar, both determiners appeared with all nouns. In the 'consistent' condition, both children and adults learned the correct determiner–noun associations. In the 'inconsistent' condition, however, adults mirrored the input distributions, using the frequent determiner 2/3 of the times, the other 1/3 of the times, without consistent associations between determiners and nouns, while children regularized the determiners, using the frequent one almost in 100% of the cases, and ignored the less frequent determiner. These findings suggest that first language acquisition is geared towards exploiting systematic relations between functors and content words.

In most of the above experiments, the markers were kept constant, while the content words varied, being drawn from categories of various sizes. Depending on the actual category size, the frequency difference between functors and content words differed from one experiment to the other. Since in natural language, functors and content words have very different type/token ratios, Gómez (2002) addressed the issue whether variability and category sizes matter for learning. She exposed her adult subjects to artificial grammar sentences that contained three words: a X c, where the a and c elements were always drawn from two 3-member sets, while the 'X' element came from a set the size of which was increased through the four conditions (2, 6, 12, 24). Thus variability increased and predictability between adjacent elements decreased across the four conditions. In addition, the grammar was created in such a way that successful learning required the recognition of the distant dependency between the a and the c elements, because 'a1' always co-occurred with 'c1' etc. Adjacent dependencies were not predictive, i.e. any a word could precede and any c word could follow any 'X' word. The results show that subjects were better at learning the distant dependency when the variability of the intervening material was high. The experiment was repeated with 18-month infants (with slightly reduced set sizes for all sets), and yielded results very similar to the adult data. These findings suggest that learners look for invariant patterns in the input.

The four previous experiments provide firm evidence that the function word/content word distinction contributes substantially to the learnability of language, at least in experimental conditions. The natural language, however, contains further information that might potentially cue syntactic structure and pave the way for learning. A natural step to take, then, is to ask how the distinction interacts with the other cues. Two such cues have been studied in greater detail, reference and prosody.

The first attempt to investigate the combined effects of markers and referents was that of Mori and Moeser (Mori & Moeser, 1983), who first replicated Green's results, and then added referents to the artificial language. They found that the presence of referents greatly facilitated learning. Moreover, when a reference field was included, subjects ignored the marker information and relied solely on the referents, even if those provided inconsistent information.

However, these results seem implausible in the face of data about blind children, who acquire their first language at the same learning speed and with the same accuracy as typically developing children despite reduced information about the world (and thus about referents) (Gleitman, 1981). Indeed, later findings by Morgan and Newport (1981); Morgan et al. (1987) and Valian and Coulson (1988); Valian and Levitt (1996) suggest that the presence of a reference field does not cancel the effects of markers, nor is it a necessary condition for learning.

Morgan and Newport (1981) investigated the conditions under which referents facilitate learning. They used differently organized reference fields with the same artificial grammar, and found that referents facilitate learning if they provide information about the constituent structure of the stream. Interestingly, no additional facilitation was obtained when referents also represented the internal hierarchical structure of the phrase. Contrary to Mori and Moeser's (1983) findings, Morgan and Newport found no facilitatory effect at all when referents mapped inconsistently onto the phrase structure of the language.

In a later study, Valian and colleagues (Valian & Coulson, 1988) combined the effects of marker frequency with reference. Their grammar contained two structures: [a A] [b B] and [b B] [a A], where a and b were markers, A and B content word categories. To manipulate marker frequency (both absolute and relative to the frequency of content words), the grammar was implemented in two different dialects. In the first dialect, markers were always lexicalized using constant tokens, while categories A and B both contained six possible content words. This dialect, then, had 2 marker tokens and 12 content word tokens, thus 14 words altogether, with a relative marker/content word frequency of 1:6. In the second dialect, each of the markers were realized by one of two tokens (obviously, different two for a and b), the content words by one of three tokens (once again, different ones for A and B). In this dialect, there were 4 marker tokens, 6 content word tokens, thus 10 words altogether, with a marker/content word frequency ratio of 2:3. In the first experiment, the authors compared the learnability of the two dialects and found that subjects learned the higher frequency (1:6) dialect faster and more accurately. In the second experiment, referents were added to the sentences in both dialects. Subjects assigned to the high frequency dialect still learned fasted and in a more accurate way than those of the low-frequency dialect. However, for both groups, learning was faster and better than in the respective groups in the previous experiment, where no reference field was given. This suggests a cumulative effect of variability and reference.

Referents are not the only cues that conspire with markers in signaling the structure of the input. Prosody is also informative, since it often indicates utterance and phrase boundaries. Morgan et al. (1987) also tested the effects of prosody on extracting the constituent structure of artificial grammars. The input sentences were generated by the same grammar as in the other two experiments of (Morgan et al., 1987). They were, then, implemented in the same three conditions as before. In the first ('no prosody'), the words were read as if they were items in a list, thus conveying no structural information. In the second condition ('inconsistent prosody'), the intonational contours were imposed on the sentences in such a way that the edges of intonational phrases were misaligned with respect to syntactic phrase boundaries. In the third condition ('consistent prosody'), prosodic boundaries were aligned with syntactic ones. The tests were similar to the ones described above: the first tested linear order and sequential co-occurrences, the second constituency.

Subjects learned the linear order and the co-occurrence patterns in all three conditions, but the ones who were exposed to consistent prosody showed a better performance in these tests, and they were the only ones who showed learning in the constituency test.

Valian and colleagues (Valian & Levitt, 1996) investigated the combined effect of markers, referents, and prosody. As a starting point, they repeated the second experiment of their previous series (Valian & Coulson, 1988) in the auditory modality. For both dialects, two prosodic realizations were created, one in which sentences were pronounced with natural phrasal intonation (rising pitch contour for the first phrase, falling for the second), and another one in which all words were read separately with a list intonation. The results replicated the previous pattern, the high-frequency grammar having been easier to learn in general. No additional effect of prosody was found. In the authors' interpretation, this shows that prosody, which is anyway a structurally less predictive cue, is not made use of when there is more reliable information available, e.g. frequency and reference. In the second experiment, therefore, the reference field was removed, while everything else was kept constant, the prediction being that in this impoverished condition, subjects would resort to prosodic information. The results confirmed the prediction. Subjects who had to learn the high-frequency dialect and who thus still had quite informative frequency cues did not benefit from prosody (no difference was found between the natural intonation and the word list intonation subgroups in this dialect), while low-frequency learners were aided by the natural phrasal intonation as opposed to the word list prosody.

These experiments strongly suggest that functors facilitate the extraction of regularities from simple artificial grammars. The next step to take is ask whether functors might play a similar role in natural language. I test this in the following two experiments.

4.2 Experiment 7: Adults show a sensitivity to the frequency distributions and word order patterns of their native language

By systematically appearing at the edges of syntactic phrases (Experiment 6 in Chapter 3), functors act as natural breakpoints bracketing the continuous input. This facilitates the extraction of regularities and the discovery of constituent structure in artificial grammars. In the current experiment, I explore another way in which the privileged syntactic positions of functors might contribute to language acquisition. The relative positions of functors and content words in the natural input correlates with basic word order (Experiment 6 Chapter 3). If this distributional information is relevant for learning word order, speakers might use it when faced with novel linguistic material, such as an artificial grammar. If this is the case, adult speakers of languages with opposite word orders, Basque, Japanese and Hungarian vs. Italian and French, should have opposite order representations, and, consequently, should organize artificial material into 'phrases' with opposite orders. The present experiment seeks to tap onto the representation of word order built through exposure to natural input in the context of learning a novel, in this case artificial, language.

4.2.1 Material

In order to obtain opposite order preferences in different populations, a structurally ambiguous artificial language was needed that allowed two contrasting organizations in terms of word order. To achieve this, a continuous stream was constructed by repeatedly concatenating a hexasyllabic basic unit aXbYcZ, where a, b, and c mimicked functors, X, Y, and Z mimicked content words (Table 4.1). The three functor categories contained one CV syllable token each, while the three content word categories comprised nine CV syllable tokens (Table 4.2). Thus, functors were identifiable through their frequency distribution, as they were nine times more frequent than content words. The hexasyllabic unit was repeated 540 times, resulting in a continuously alternating sequence of functors and content words. This stream was rendered ambiguous by eliminating initial and final phase information (ramping the first 15 sec of the stream up and the last 15 seconds down in amplitude). This way, the basic unit could be perceived as starting with a frequent functor (henceforth: frequent-initial items: aXbYcZ, e.g. /fibanutagebi/, cZaXbY, e.g. /gekufipanufe/ etc.) or with a non-frequent content word (henceforth frequent-final items: YcZaXb, e.g. /kɔgenapifenu/, XbYcZa, e.g. /tɔnukɔgeʀifi/ etc.). This ambiguous stream was used for familiarization. Test items were hexasyllabic. Eighteen instantiated the frequent-initial order, another 18 the frequent-final order. All three functor and content word categories were used with equal frequency in each position within test items. Content word tokens were used four times each to make up the 36 different test items.

Structure	...**a** X **b** Y **c** Z **a** X **b** Y **c** Z **a** X **b** Y **c** Z **a** X **b** Y **c** Z...
Stream	...**fi**lu**nu**fe**ge**mu**fi**pe**nu**ta**ge**li**fi**du**nu**pi**ge**ʀɔ**fi**pa**nu**mɔ**ge**bi...
Ambiguity	...**fi**lu**nu**fe**ge**mu**fi**pe**nu**ta**ge**li**fi**du**nu**pi**ge**ʀɔ**fi**pa**nu**mɔ**ge**bi... OR ...**fi**lu**nu**fe**ge**mu**fi**pe**nu**ta**ge**li**fi**du**nu**pi**ge**ʀɔ**fi**pa**nu**mɔ**ge**bi...

Table 4.1: The familiarization stream used in Experiment 7.

When creating the lexicon, care was taken to avoid phonotactic biases. All 36 syllables were non-words and had similar frequencies in word initial positions in the five languages.

The familiarization stream was synthesized with the es1 (Spanish male) voice of the MBROLA software (Dutoit, 1997). The es1 voice was selected after pilot studies with six different MBROLA voices (de4, de6, es1, es2, it1, it2) testing discriminability. Phonemes were all 120 msec long and had a

a	X	b	Y	c	Z
/fi/	/ʀu/	/nu/	/fe/	/ge/	/mu/
	/pe/		/ta/		/ʀi/
	/du/		/pi/		/ku/
	/ba/		/be/		/bɔ/
	/fɔ/		/bu/		/bi/
	/de/		/kɔ/		/dɔ/
	/pa/		/mɔ/		/ka/
	/ʀa/		/pɔ/		/na/
	/tɔ/		/pu/		/ʀɔ/

Table 4.2: The lexicon of CV words used in Experiment 7.

uniform 100Hz pitch, making up a non-intonated, monotonous stream, which was about 17 minutes 30 seconds long. Test items had the same parameters, and were 1440 msec long.

4.2.2 Languages

Languages were chosen to represent different typological options, and crucially, opposite basic word orders. In Italian (for a more detailed description, see Chapter 3, Section 3.1.2) Specifiers precede Heads and Heads precede their Complements. French has the same properties, and was selected to confirm and replicate the results obtained with Italian. In these languages, most functors appear at the left edges of phrases (see also Chapter 3). Therefore, they were expected to give rise to frequent-initial word order preferences.

Hungarian (see Chapter 3, Section 3.1.2) is characterized by Specifier-Head order and mixed, but predominantly Complement-Head basic word order. It might be expected, therefore, that subjects will show no clear preference, or a weak bias towards a frequent-final order.

Japanese (see Chapter 3, Section 3.1.2) is a Specifier-Head and Complement-Head language. Like Japanese, Basque also shows Complement-Head order, but unlike the other four languages, Specifiers follow Heads. In these two languages, most functors are to the right. Consequently, subjects were expected to prefer frequent-final orders. This preference could be stronger in

Basque than in Japanese due to the Head-Spec order in the former language.

4.2.3 Subjects

All subjects were recruited on a voluntary basis, were naive with respect to the purpose of the experiment, and were paid for their participation. Care was taken to select subjects who had little or no knowledge of languages other than their native language.

Basque subjects

Twelve adult native speakers (approx. half were females; mean age: 27 years, range: 20-37) of Basque participated in the experiment. Due to sociopolitical circumstances, Basque native speakers also speak Spanish. However, care was taken to select subjects (i) who were late ($\geq$ 2 years of age) learners of Spanish, (ii) whose parents were native or native-like speakers of Basque, (iii) who use Basque in their daily interactions with family and friends, and (iv) who live in Basque-speaking areas.

Japanese subjects

Twenty-four adult native speakers (18 females; mean age: 22 years, range: 20-28) of Japanese participated in the experiment.

Hungarian subjects

Thirty-three adult native speakers (6 females; mean age: 21 years, range: 19-27) of Hungarian participated in the experiment.

Italian subjects

Twenty-nine adult native speakers (16 females; mean age: 24 years, range: 20-34) of Italian participated in the experiment.

French subjects

Twenty-one adult native speakers (13 females; mean age: 23 years, range: 19-29) of French participated in the experiment.

4.2.4 Procedure

Subjects were tested individually in sound-attenuated booths or silent rooms depending on the location. They were seated in front of a computer screen, where the instructions appeared. Sound stimuli were administered to them through high quality headphones.

At the beginning of the experiment, subjects were instructed that they would listen to a sample of an unknown language, and would then be tested on their knowledge of the 'sentences' of the language. A short training session followed in order to familiarize subjects with the two-alternative forced choice procedure, used later in the test phase. During this training, subjects heard 10 syllable pairs. In each pair, they had to identify a target syllable by pressing one of two predefined keys depending on whether the target syllable appeared as the first or the second item of the pair. After training, subjects were instructed to listen to the familiarization stream, which lasted 17 minutes 30 seconds. After familiarization, subjects passed immediately onto the test phase. In each of the 36 trials, they heard a pair of 'sentences', and they had to indicate by pressing one of the two predefined keys which of the two 'sentences' sounded more like a possible sentence of the unknown language. Items within a pair were separated by a pause of 500 msec.

Each test item was tested against another one that represented the opposite word order. All test items were used twice, once as the first member of a test pair and another time as the second. The same test item never appeared in consecutive trials. The order of presentation was randomized and counterbalanced across subjects.

Basque subjects were tested in the Vitoria-Gasteiz area (Basque Country/Spain) by members of the Cognitive Neuroscience Research Group, Department of Psychology, University of Barcelona. Japanese subjects were tested at the Laboratory of Language Development of the Brain Science Institute of RIKEN, Tokyo, Japan. Hungarian subjects were tested at the Department of Cognitive Sciences, Budapest University of Technology and Economics, Budapest, Hungary. Italian subjects were tested at the Language, Cognition and Development Laboratory of SISSA, Trieste, Italy. French subjects were tested at the Cognitive Sciences and Psycholinguistics Laboratory, EHESS/ENS/CNRS, Paris, France.

4.2.5 Results

The number of frequent-final responses was registered (Figure 4.1) and entered into data analysis. (The number of frequent-initial responses can be obtained by subtracting this number from the total number of trials (36). An ANOVA with the factor Word Order (Comp-Head / Mixed / Head-Comp) yielded a significant main effect ($F(2, 116) = 10.554, p \geq 0.0001$). For further pair-wise comparisons between word order types, Bonferroni post hoc tests were conducted. The clear Complement-Head languages (22.81 ± 4.93) did not differ significantly ($p = 0.99$) from the mixed Complement-Head language Hungarian (23.21 ± 6.60), but differed significantly ($p = 0.0007$) from the Head-Complement languages (17.20 ± 7.99). These latter also differed significantly ($p = 0.0004$) from the mixed Complement-Head language.

When a similar ANOVA was conducted with the factor Language (Basque / Japanese / Hungarian / Italian / French), a significant main effect was obtained ($F(4, 114) = 6.193, p = 0.0002$). Languages were compared pairwise in a Bonferroni post hoc test. Italian differed significantly from Basque ($p = 0.004$) and Hungarian ($p = 0.007$), and so did French ($p = 0.007$ and $p = 0.018$, respectively). No other comparisons were significant.

The order preference scores of the Basque, Japanese and Hungarian groups were significantly above chance at $\alpha = 0.01$, corrected for multiple comparisons ($t(11) = 5,541, p \geq 0.0001; t(23) = 3.754, p = 0.001; t(32) = 4.534, p \geq 0.0001$; respectively). The preference scores of the Italian and French groups were below chance, but the difference did not reach statistical significance ($t(21) = -0.457$, ns.; $t(29) = -0.636$, ns.; respectively).

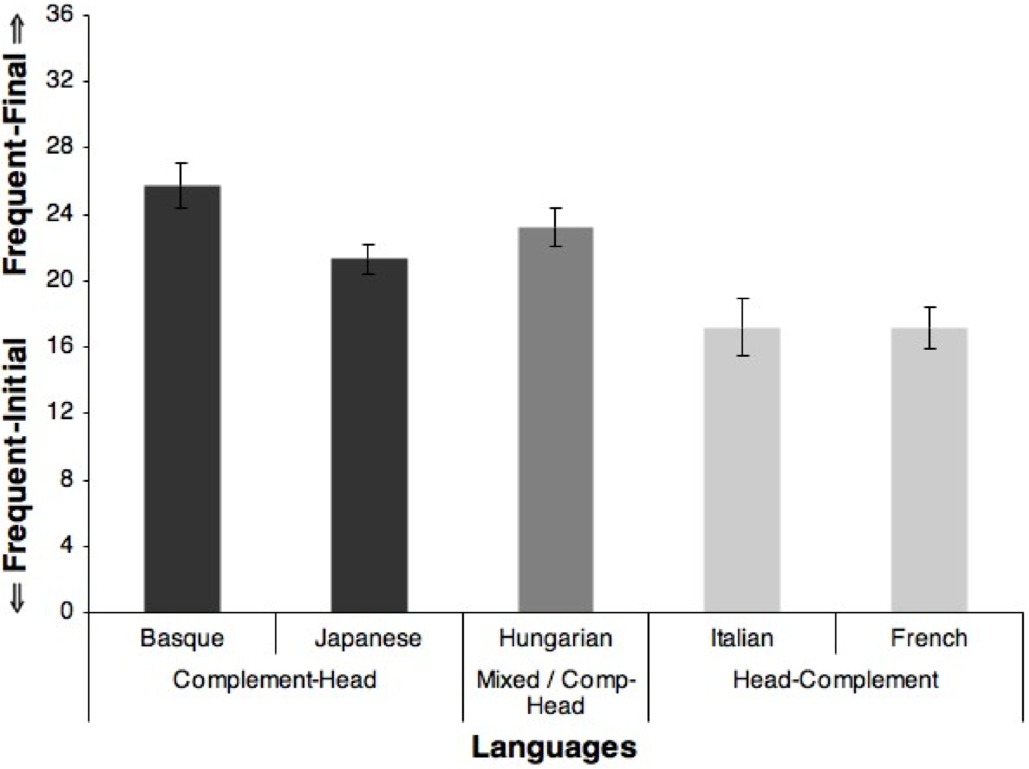


Figure 4.1: Preference scores for frequent-final word order in Experiment 7. The x-axis represents the five languages, grouped and color-coded according to their word order type. The y-axis shows the number of frequent-final responses. Error bars represent the standard errors of the means.

4.2.6 Discussion

Adult native speakers of five languages representing three different word order types have been found to show different order representations, corresponding to those found in their respective native languages. Specifically, speakers of Basque, Japanese and Hungarian, i.e. the Complement-Head languages, exhibited a preference for phrases ending in frequent elements, mimicking functors. It appears that the mixed Complement-Head language, Hungarian, is not different from the clear cases in this regard. This is not surprising, since mixed orders are only attested in verb phrases; other phrase types uniformly show Complement-Head order. The Basque group, as predicted, did show a somewhat higher preference than the other two Complement-Head languages, probably due to its Head-Specifier property, but this difference was not significant, indicating the relative strength of the Complement-Head property.

The Complement-Head speakers' preference differed from the Italian and French speakers' responses, which, contrary to predictions, exhibited no significant preference (although their direction was as expected). The absence of an effect in this group requires further clarifications. However, one possible explanation might be that adults, who are not learners, but mature speakers, have sophisticated representations (e.g. in terms of parameters) of order phenomena present in many different components of their native language. Consequently, the frequency distributions characteristic of components other than syntax, e.g. morphology, might have influenced their choices. While the Complement-Head languages in my sample have suffixing morphology (with some verbal prefixing in Hungarian), which converges with the frequent-final bias of word order. The Head-Complement languages included in the sample are inflecting. This makes no specific contribution to the order preference, since in these morphological systems, it is the stem itself that changes, no functional morpheme is added. However, both Italian and French has some (mostly derivational) suffixing. This goes against the frequent-initial bias deriving from word order properties. As Austin et al. (n.d.) have shown, adults tend to mirror frequency distributions in the input very closely. Therefore, it is not implausible to assume that subjects were influenced by different distributions of their native languages, modulating their preference.

4.3 Experiment 8: Infants show a sensitivity to the frequency distributions and word order patterns of their native language

Since infants make use of distributional information in a more categorical manner (Austin et al., n.d.), have less knowledge of the frequency patterns of the target language, and, importantly, are the primary users of bootstrapping strategies, the hypothesis of a frequency-based bootstrapping mechanism to cue word order needs to be tested in this population.

The existence of such a word order representation can also shed light on a long-standing theoretical debate about the acquisition of word order, and syntax in general. As discussed in Chapter 1, lexicon-based, constructivist theories of syntax acquisition (Tomasello, 2000) argue that word order is not acquired as a general, abstract structural property. Rather, order patterns are learned separately for each lexical item. Consequently, word order cannot be learned *before* a small initial lexicon is built. Under the generativist view, acquiring word order is a matter of setting the relevant parameters. Therefore, it is independent of the lexicon. The litmus test to decide between the two theories, then, is to investigate whether any general knowledge of word order exists prior to the acquisition of at least some basic lexicon.

In the current Experiment, therefore, I tested the word order preferences of prelexical (8-month-old) Japanese and Italian infants in the headturn preference paradigm. Given the technical exigencies of recruiting and testing infants, only two of the previous five populations participated in this study. They were chosen in such a way as to represent the two critical word order types, for which frequency cues have been shown to exist in the input (Chapter 3). If opposite preferences are obtained in the two populations, that provides evidence for the existence of a rudimentary word order representation, bootstrapped through frequency distributions in the input, and argues for the generativist acquisition model.

4.3.1 Material

The material of the adult Experiment was adapted and simplified for the purposes of this Experiment. Thus a tetrasyllabic basic unit was used: aXbY, where, as before, a and b represent frequent functors with one token in each category, while X and Y are content word categories containing 9 tokens each (Table 4.3). Similarly to the adult material, phase information was suppressed, by ramping the amplitude of the initial and final 15 sec of the stream. The four-syllabic basic unit was repeated 243 times (each possible _X_Y syllable combination was used 3 times), resulting in a 3 min 53 sec long familiarization stream. The CV words used in the functor and content word categories were also adapted from Experiment 7.

Structure	...**a** X **b** Y **a** X **b** Y **a** X **b** Y **a** X **b** Y...
Stream	...**fi**lu**ge**mu**fi**pe**ge**li**fi**du**ge**ʀɔ**fi**pa**ge**bi...
Ambiguity	...**fi**lu**ge**mu**fi**pe**ge**li**fi**du**ge**ʀɔ**fi**pa**ge**bi... OR ...**fi**lu**ge**mu**fi**pe**ge**li**fi**du**ge**ʀɔ**fi**pa**ge**bi...

Table 4.3: The familiarization stream used in Experiment 8.

The familiarization stream was synthesized using the fr4 female diphone database of MBROLA (Dutoit, 1997), with a monotonous pitch of 200Hz and a constant phoneme duration of 120 sec. These modifications were necessary to render the acoustic properties of the material more pleasant and interesting for infants.

Test items were 8 four-syllabic ‘sentences’ of the language, 4 instantiating the frequent-initial order (aXbY: e.g. /gemufide/), the other 4 the frequent-

a	X	b	Y
/fi/	/ʀu/	/ge/	/mu/
	/pe/		/ʀi/
	/du/		/ku/
	/ba/		/bɔ/
	/fɔ/		/bi/
	/de/		/dɔ/
	/pa/		/ka/
	/ʀa/		/na/
	/tɔ/		/ʀɔ/

Table 4.4: The lexicon of CV words used in Experiment 8.

final one (XbYa: e.g. /dugeʀifi/). The X and Y words making up the test items were chosen in such a way that the transitional probabilities (TP) between all syllable pairs within test items be zero or very low both in Japanese and Italian, as measured in the respective corpora of infant-directed speech used in Experiments 5 and 6. In an ANOVA with factors Language (Japanese/Italian) X Order (frequent-initial/frequent-final), using as dependent measure the TPs between the syllable pairs contained in the test items, I found no significant main effect (Language: $F(1,44) = 0.13, p = 0.72$; Order: $F(1,44) = 0.54, p = 0.46$) or interaction ($F(1,44) = 1.39, p = 0.24$). Thus, there was no bias in the test items from the TPs of the native language.

A test trial consisted of 15 repetitions of the same test item, separated by 500 msec pauses. The order and side of presentation of the test trials was randomized and counter-balanced across subjects in such a way that at most two consecutive trials could be of the same order type (frequent-initial/frequent-final).

4.3.2 Subjects

The Japanese group consisted of 20 8-month-old infants (9 females; mean age: 235 days, age range: 201254 days). They were born to monolingual Japanese families, and had no record of neurological or auditory impairment. An additional 11 babies were tested, but not included in the analysis for the following reasons: failure to complete the experiment due to crying (3), fussiness (7), and experimenter error (1).

The Italian group consisted of 20 8-month-old infants (10 females; mean age: 234 days, age range: 214256 days). They were born to monolingual Italian families, and had no record of neurological or auditory impairment. An additional 10 babies were tested, but not included in the analysis for the following reasons: failure to complete the experiment due to crying (4), fussiness (4), experimenter error (1), and technical error (1).

A parent of each infant gave informed consent prior to participation. The study was approved by the Ethics Committee of RIKEN (where the Japanese infants were tested) and the Ethics Committee of SISSA (where the Italian infants were tested).

4.3.3 Procedure

I used a version of the headturn preference paradigm as described in Saffran, Johnson, Aslin, and Newport (1999) to test infants' word order preferences. Infants were tested individually while sitting on a parents lap in a dimly lit, sound-attenuated cubicle (Figure 4.2). Parents were listening to masking music and were wearing dark sunglasses throughout the experiment to avoid all parental influence on infants' behavior. Infants first listened to the almost 4-minute-long familiarization stream, while they watched attention-getter lights at the two sides or the center of the testing cubicle. The blinking of the lights was contingent upon the infants' looking behavior, but there was no systematic relation between the lights and the sounds. During the experiment, an experimenter, blind to the stimuli and seated outside the testing cubicle, monitored infants' looking behavior and controlled the lights and the stimuli. Infants were videotaped during the experiment for the subsequent off-line coding of their looking behavior.

Immediately after familiarization, infants were tested for their word order preference in 8 test trials. Each trial started with the blinking of the central light to attract infants' attention. Once infants attended to the central light, one of the side lights started blinking and the central light was extinguished.

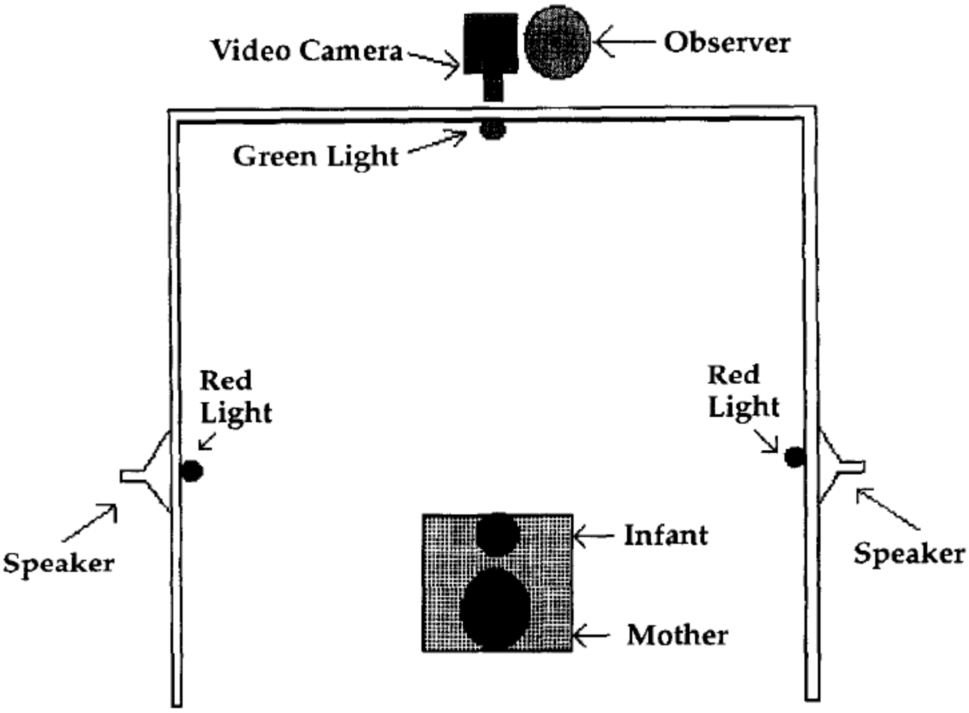


Figure 4.2: The experimental setup of the headturn preference paradigm used in Experiment 8. Image adapted from Kemler Nelson et al. (1995).

When infants stably fixated on the blinking side light (defined as a 30 ° head turn towards the light), the associated test item started playing from a loudspeaker on the corresponding side. The sound file continued until the end (22 sec) or until infants looked away for more than 2 sec. After this, a new trial began.

Japanese infants were tested at the Laboratory of Language Development of the Brain Science Institute of RIKEN, Tokyo, Japan. In this laboratory, the testing cubicle had real lamps mounted on its walls. Italian infants were tested at the Language, Cognition and Development Laboratory of SISSA, Trieste, Italy. Here, the attention-getter lights were implemented as movies of blinking lamps displayed on flat screens attached to the walls of the cubicle.

4.3.4 Results

Infants' looking times were coded and measured off-line. They were averaged across all trials of the same type (frequent-initial/frequent-final) for both groups (Figure 4.3). An ANOVA with factors Language (Japanese/Italian) as a between subject variable and Order (frequent-initial/frequent-final) as a within subject variable, using looking times as the dependent measure, yielded no significant main result. Importantly, the interaction Language X Order was significant ($F(1, 38) = 8.3301, p = 0.006$), indicating that the two groups showed opposite looking patterns.

In a Scheffe post hoc test, I also compared looking times for the two types of test items in each group. The Japanese group looked significantly longer at the frequent-final items over the frequent-initial ones ($p = 0.046$), whereas the Japanese group exhibited the opposite pattern ($p = 0.049$).

4.3.5 Discussion

These results show that Japanese and Italian infants, who are exposed to languages with opposite word orders, have opposite expectations about the order of frequent and infrequent items in their target language. In other words, they show sensitivity to the frequency distributions and words orders they encounter in the input. This suggests that infants might use the relative

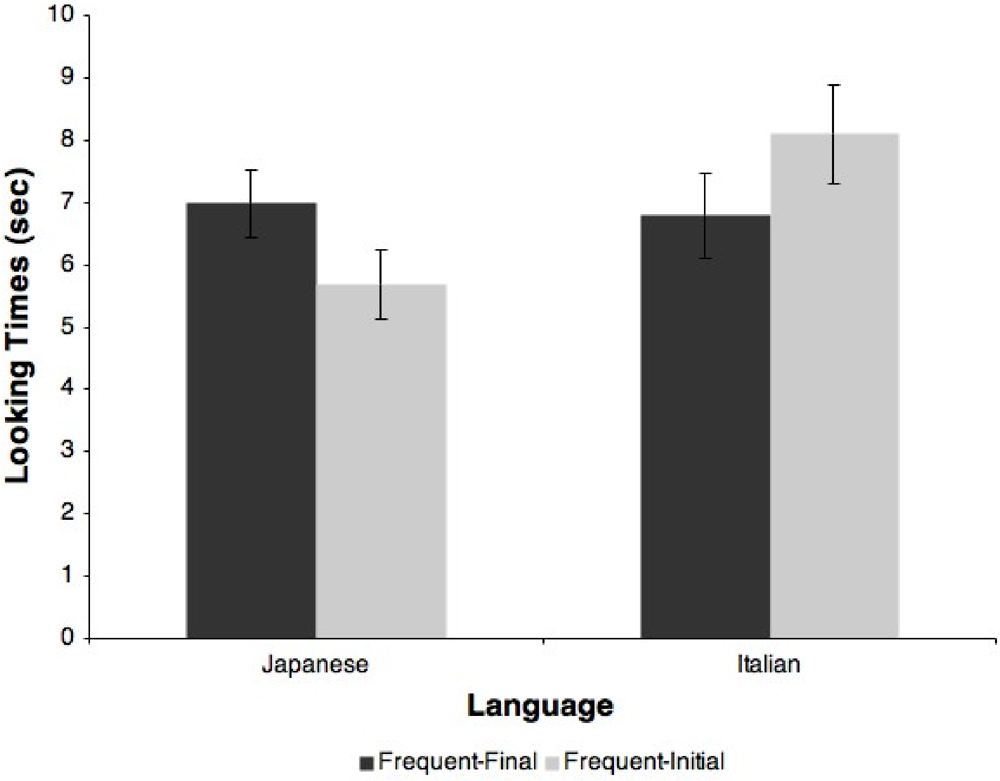


Figure 4.3: The average looking time values of the Japanese and Italian infants in Experiment 8. The x-axis represents the two types of test items in the two groups. The y-axis indicates looking times in seconds. Light grey bars represent average looking times to the frequent-final test items. Dark grey bars represent average looking times for the frequent-initial test items. Error bars show standard errors of the mean.

order of functors and content words at utterance boundaries to create one of the first rudimentary representations of word order already before they build their lexicon. Therefore, these results support the frequency-based bootstrapping hypothesis, and the generativist view of acquisition in general.

Note that Italian babies, unlike Italian adults, showed a preference significantly different from chance. This provides some indication that infants might indeed 'regularize' the distributions encountered in the input, as Austin et al.'s (n.d.) studies with children suggest. This is a useful strategy if they are to extract general regularities rather than learn particular details from the input.

4.4 General Discussion: Bootstrapping word order from early frequency-based representations

The two experiments presented above asked whether infant learners and adult speakers of different languages have some abstract representation of the basic word order of their language in terms of the relative positions of functors and content words. This was tested by assessing subjects' word order preferences when breaking into the structure of an ambiguous artificial speech stream, consisting of a regular alternation of frequent 'functor' elements and infrequent 'content word' elements. Opposite preferences, corresponding to the distributions of the native language (as shown in Chapter 3, Experiment 6), have been found in the populations learning/speaking languages with opposite word orders.

These preferences have been more pronounced in infants. Further research will be necessary to explore this difference. One possibility, however, supported by previous findings (Newport, 1990; Austin et al., n.d.), is that infants, engaged in the task of acquisition, use different mechanisms than adults, allowing them to extract the maximum amount of information from the relatively small amount of exposure they have received. These mechanisms might be predisposed to project generalizations of the basis of limited input (the “Less is More” principle proposed by Newport, 1990, for experimental evidence with 8- and 13-month-old infants, see Marchetto & Bonatti, in preparation), and as a corollary, to regularize, i.e. to categorically represent statistical input (Austin et al., n.d.).

In the light of the above, I propose, then, that the relative order of functors and content words, signaled in the input as the relative position of frequent and infrequent elements, especially at utterance boundaries, might cue the basic word order of the target language. Young learners use this to prelexically establish an initial representation of word order, which can then serve, most probably in conjunction with other bootstrapping mechanisms, to break the input down into its rough syntactic constituents.

The results of Gomez (Gómez, 2002) also contribute very importantly to the issue of frequency. Her conclusion is that high frequency differences between the two categories make markers, and the dependencies between them more salient. In a certain sense, the high frequency difference observed in natural language makes it possible for the learners to zoom in on the invariant grammatical skeleton of the sentence and on the relations that hold among its elements. When such invariant structure cannot be found, language learning can become greatly disturbed. The high frequency of function words and their resulting perceptual saliency offers a way out from the logical paradox of distributional analyses pointed out e.g. by (Pinker, 1984; Cartwright & Brent, 1997), namely that in order to derive categories from their distributions, the neighboring categories need to be known, but categories are precisely what has to be learnt. The learner thus needs non-distributional cues to first break into the system. This is exactly what high frequency can provide. This is further evidence for the functor/content word distinction as a necessary design feature of natural language, since the frequently recurring functors provide precisely the required invariant structure in the sea of highly variant content words.

This frequency-based bootstrapping mechanism raises a number of issues that need further discussion. First, what level of abstraction is the representation encoded at? Does it serve as a trigger to set the abstract word order parameters? Or does it remain a statistical encoding? While these experiments provide no definitive answer, it can safely be concluded that the representation is abstract enough to allow infants to generalize it onto the functors and content words of an unknown artificial language, in which all the words are novel for them. However, more speculatively, I propose that the frequency-based order representation works in concert with other bootstrapping mechanisms such as prosodic bootstrapping (Nespor et al., 1996, under review) to establish a more fully-fledged representation of word order in the target language. I suggest that the frequency-based mechanism works as an initial, universal procedure, yielding a general, overall representation of the most dominant word order pattern, characteristic of most phrase types in the target language. Then, this initial representation might be further elaborated by prosodic bootstrapping mechanisms, which assign a precise word order to each phrase type, especially when the word order of a given phrase type is different from the dominant order of the language. For instance, the prosodic bootstrapping mechanism (Nespor et al., 1996, introduced in Chapter 1), establishing word order on the basis of the position and the physical realization of prosodic prominence in phonological phrases, might serve precisely this function. Importantly, this mechanism, just like the frequency-based one, allows bootstrapping word order independently of the lexicon, and it also makes use of the edge positions of phrases. Given these representational similarities, it is not implausible that the two mechanisms might complement each other during the acquisition of word order. This hypothesis is in line with other bootstrapping theories that emphasize the importance of convergent cues in language acquisition (Morgan & Demuth, 1996).

A second issue concerns the phonetic form and detail in which the frequency-based representation might be encoded. We tested 8-month-old infants, who are clearly prelexical. From work by Shi, Cutler, et al. (2006); Shi, Werker, and Cutler (2006), we know that at this age, infants have an underspecified representation of functors. Such an underspecified representation, however, does not compromise our hypothesis, since word forms are simply required to be categorized as a functor, their unique identification is not necessary. Indeed, (Redington et al., 1998; Mintz et al., 2002) have shown that in corpus-based simulations of categorization, replacing collapsing functors into one general functor category did not seriously compromise performance. The fact that 8-month-olds only track the most frequent functors is also consistent with my proposal, since, as I have show (Chapter 3, Experiment 5), the distributions of functors and content words are maximally distinct precisely in the highest frequency range. As we have observed, a handful of the most frequent functors already cover a large part of the input, and provide reliable information about word order.

From a methodological point of view, the Experiments in this Chapter constitute an effort to go beyond simple artificial grammars in the study of language acquisition. Artificial grammars have been very useful tools in experimental psycholinguistics since the seminal works of Reber (1967). By allowing to control for and manipulate experimental factors in a systematic way, they have been instrumental in isolating and identifying different computational abilities both in the study of language processing and language learning/acquisition (rule learning: Gómez & Gerken, 1999; Marcus et al., 1999; statistical learning: Saffran, Aslin, & Newport, 1996 etc.). However, it has been recently recognized and emphasized (Morgan et al., 1996) that single abilities or single cues are very often not sufficient to explain acquisition, since linguistic phenomena are themselves complex and are characterized by a group of correlated cues at different levels of description (this will be amply illustrated below for the functor/content word distinction). Therefore, recent studies in acquisition (Chambers, Onishi, & Fisher, 2003; Graf Estes, Evans, Alibali, & Saffran, 2007) strive to combine artificial grammars with natural language. The experiments presented below achieve this by bringing prior knowledge in natural language to bear on an artificial grammar learning task.

Chapter 5

General Discussion and Open Questions

The three previous Chapters have sought to explore how three mechanisms—perceptual primitives, statistical learning and rule extraction—contribute to young infants' acquisition of their mother tongue. Below, I will review each of them separately, consider what evidence has been found concerning their respective role in language acquisition, and discuss some issues and open questions that arise. Then, I will consider the interactions that the experiments have revealed between these mechanisms, and will outline an approach of language acquisition in which the mechanisms act in concert, proving input for and placing constraints on each other.

5.1 Three basic mechanisms of language acquisition: perceptual primitives, statistics and rules

5.1.1 Perceptual primitives in the initial state

Perceptual primitives have been operationally defined in Chapter 1 as feature or object configurations that honor the sensitivities and stimulus preferences of a given sensory system and are thus automatically and efficiently detected.

At least two such configurations were identified for the auditory system in adults: repetitions and edges (Endress et al., 2005, in press). However, it was unknown whether auditory perceptual primitives are available to infants and whether they play a role in language acquisition. Since hearing begins around the 20th week of gestation and the auditory system reaches a certain level of maturity by the 20th–24th week (Mehler & Dupoux, 1994; Moore & Jeffrey, 1994; Moore, 2002), it is expected that some adult characteristics of auditory perception are present already at birth.

Therefore, in Experiments 1–3, I have investigated whether repetitions might function as perceptual primitives in the neonate auditory system. Indeed, I have found that the newborn brain reacts to repetitions with significantly increased activation in the temporal areas and significantly increased or decreased activation in the frontal areas (depending on the specific structural properties of the stimuli), as compared to random controls. However, it has also been observed that not all repetition configurations are perceptual primitives. Adjacent repetitions (ABB and A_A; Experiments 1 and 3) are distinguished from their controls, whereas non-adjacent repetitions (ABA; Experiment 2) are not.

These results raise certain questions with respect to language acquisition. Can such a limited identity-detector help infants in the task of exploring their linguistic environment? At what level of abstraction can the identity-detector operate? Does this change throughout development? If yes, when and how? Are these changes related to the ability of recognizing identity at a distance? I will address these questions in turn.

The experiments in Chapter 2 used artificial grammars to establish that the newborn brain detects immediate repetitions. But what is the role of this mechanism in the acquisition of a natural language? Several possible applications exist. First, infant-directed speech is rich in immediate repetitions (Sundberg, 1998). When talking to children, adults very often repeat words or even whole phrases identically. This might facilitate processing for infants. More importantly, typical 'child words' often contain full or partial reduplications in many languages (e.g. *baby*, *bébé* [French], *baba* [Hungarian]; *daddy*, *papà* [Italian]; *dodo* 'sleep' [French], *csicsika* 'sleep' [Hungarian] etc.). These are also very often the child's first words. Thus, the perceptual saliency of these words might help infants discover and learn them as the first entries in their lexicons.

As discussed before, however, infant-directed speech might facilitate the task of young learners, but it is by no means necessary for the success of language acquisition. So the functions of reduplication in adult-directed speech also needs to be considered. Reduplication is an operation characteristic of morphology. In derivational morphology, its most common function is the formation of onomatopoeic words in language that make extensive use of this vocabulary stratum (e.g. /pikapika/ 'to blink' in Japanese). In inflectional morphology, it is most typically used in the formation of diminutives (e.g. Bikol[1]: *aloy* 'time span', *aloy-aloy* 'short time'), plurals (e.g. Bikol: *bulan* 'month', *bulan-bulan* 'every month') or verbal aspect (e.g. Tagalog: *bili* 'buy', *bibili* 'will buy'). It is interesting to note in this regard that reduplication, as many morphological operations, applies to the edges of word stems. As discussed earlier in Chapters 1 and 2, edges are also perceptual primitives. Thus, it seems that repetitions and edges typically converge in natural language morphology—just as they did in the artificial grammar stimuli used in the experiments. Since we know that suffixing is much more common in natural languages than prefixing (Julien, 2002), it will be interesting in the future to experimentally compare repetitions that appear at the left vs. the right edge of sequences (e.g. ABB vs. AAB).

The natural language phenomena mentioned so far all concerned repetitions identical down to the level of the phoneme, as did the experiments themselves. Yet, adults are able to recognize identity at more abstract levels. They are able to recognize that the monologue of Hamlet recited by Lawrence Olivier or Peter O'Toole is actually the same text its many differences notwithstanding. Adults are able to identify a person even when she changes cloths or hairstyles, or when she is seen from a different angle. In natural language, abstract identity also plays a crucial role. *John* and *the book that you asked me whether I have ever seen*, for instance, are both DPs[2]

[1] Bikol is a Central Philippine language from Southern Luzon.

[2] For abbreviations, see p. 1.

at a certain level of abstraction, and it is this identity that allows them to have the same functions in a sentence, e.g. the Subject. Clearly, identity relations of this type are not detected by a perceptually-based identity-detector. But since repetitions at the phonemic level already represent abstractions with respect to the acoustic signal, the question arises at what level of granularity the perceptual identity-detector operates and how it relates to the representation of abstract identity. These issues will need to be determined by future experiments, e.g. those that manipulate different physical aspects of the stimuli such as speaker identity, pitch, duration etc.

A related question is why non-adjacent repetitions fail to activate the identity detector. Is it due to a limitation of the neonate perceptual system, which later disappears with development? Or is it because adjacent and non-adjacent repetitions are genuinely different phenomena even for the adult perceptual system; i.e. adjacent repetitions are perceptual Gestalts and non-adjacent repetitions are not? If this latter is the case, then how are non-adjacent repetitions detected? Do they require some abstract, symbolic representation? If so, is it the same kind of abstract identity recognition as described above? Some of these questions may be answered by testing older infants and adults in brain imaging studies using different repetition-based stimuli.

Although many questions remain open, it has been established that newborn babies share some of the perceptual Gestalts that adults and other animals are also sensitive to. These perceptual biases might assist infants in establishing some of the first word candidates and in exploring certain morphological regularities of natural language.

5.1.2 Statistical learning

Learning driven by the statistical properties of the linguistic signal have often been invoked to explain segmentation, word learning (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996; Saffran et al., 1999), lexical categorization (Mintz, 2003; Mintz et al., 2002) and even the acquisition of certain syntactic operations. Studies have also investigated the distributions of statistical cues in the input, showing that such measures reliably signal word boundaries (Batchelder, 2002; Brent & Cartwright, 1996; Swingley, 2005; Yang, 2004) and lexical categories (Mintz, 2003; Mintz et al., 2002). However, these studies mainly looked at English and a few other, typologically similar languages. Therefore, in Chapter 3, I have computed a number of statistical measures over infant-directed corpora of Hungarian, Italian and Japanese, and have shown how they cue the segmentation and categorization of lexical items, as well as word order. Below, I will summarize the results obtained for conditional probabilities and frequency separately.

Conditional probabilities

Comparing segmentation in an agglutinating and an inflecting language, I found that forward TPs, backward TPs, their combined use and mutual information all indicate word boundaries with relatively high accuracy in both languages. However, the most accurate predictor differed as a function of the morphosyntactic properties of the languages. In Hungarian, which is agglutinating and Complement-Head, BWTPs have proven to be more efficient, whereas in Italian, an inflecting and Head-Complement language, the best results were obtained for FWTPs.

This result suggests that conditional probabilities, in addition to segmentation, might also be informative about general morphosyntactic properties. While this is an intriguing possibility, further investigations are needed to clarify several issues. First, while language-specific differences have been found between the accuracy of transitional probabilities, their overall distributions are very similar both within and across languages. Therefore, it is not clear what aspects of TP computations, if any, might act as bootstrapping cues. Second, while it has been amply demonstrated that infants, adults and animals are able to compute FWTPs, it is not known whether they can also compute BWTPs or any backward going statistical information, and whether they can compare the informativeness of conditional probabilities of different directionalities. Experimental evidence is required to clarify these points.

In addition to the usual word-level segmentation, the experiments also addressed the question whether conditional probabilities signal affix boundaries, facilitating the acquisition of agglutinating morphology. It has been found that when the segmentation algorithm generates one type of boundary (i.e. boundaries are posited only at the points of the lowest coherence), these match actual word boundaries and actual morpheme boundaries with roughly the same accuracy. However, when the segmentation algorithm generated two types of boundaries (word boundaries at the lowest values of coherence and word internal morpheme boundaries at intermediate values of coherence), the resulting word boundaries coincided well with actual word boundaries, while morpheme boundaries were recognized less accurately. The reason for the less accurate performance on morpheme boundaries was the high number of false alarms, i.e. cases where the algorithm predicted a morpheme boundary, but the corpus actually had a word internal transition.

The finding that algorithms based on asymmetric conditional probabilities tend to overgenerate word internal morpheme boundaries might seem surprising at first. When considered in the larger context of language acquisition, however, some interesting implications follow. In morphologically complex languages, the learner has a double task of efficiently decomposing polymorphemic word forms into stems and affixes in such a way that (i) word candidates can be established for both stems, i.e content words, and affixes, i.e. functors; and (ii) the correct combinatorial regularities of the morphological system be learned. However, the task of extracting the correct word and affix candidates is hindered by the fact that agglutination is typically accompanied by a number of morpho-phonological processes that ensure the coherence of the complex word form (Bickel & Nichols, 2007). For instance, many agglutinating languages like Hungarian or Turkish have vowel harmony, whereby the vowels of the affixes change according to the vowel set used by the stem (e.g. Hungarian *a ház-b**a**n* the house.INE ‘in the house’, but *a kéz-b**e**n* the hand.INE ‘in the hand’). Agglutinating languages usually also apply assimilation processes or other processes creating stem or affix allomorphy when stems and affixes are concatenated. The Hungarian instrumental case marker *-val/-vel*, for example, in addition to harmonizing, also assimilates its initial consonant to the final consonant of the stem (e.g. *kéz-**z**el* hand.INSTR 'by hand'; *láb-**b**al* foot.INSTR 'by/with the foot'; *az almá-**v**al* the apple.INSTR 'with the apple'). Moreover, agglutinating languages often have word-level stress at a fixed position within the word. In Hungarian, stress is always on the first syllable of a word, while in Turkish, it is word-final.[3] As an effect of these converging mechanisms, the morphologically complex word form is a coherent morpho-phonological domain, which renders the extraction of its constituent morphemes more difficult by obscuring their boundaries.

If statistical computations and morpho-phonological processes work in concert, one important function of the former might be to help decompose the morphemes within the morpho-phonological domain of the complex word. In fact, similar interactions between statistical segmentation and phonological domains have been observed at the level of the intonational phrase. Shukla et al. (2007) have observed that in a continuous artificial speech stream that has prosodic contours overlaid on it, participants preferred statistically coherent words that were placed inside the prosodic contour to statistically similarly coherent words that spanned prosodic boundaries. This preference was the result of a mapping or alignment process between the output of statistical computations and the prosodic bracketing of the stream.

The proposal that statistics might serve to segment polymorphemic word forms is convergent with psycholinguistic studies of morphological processing in Hungarian adults. Pléh and Juhász (1995) found that prefixes and case suffixes are stripped off the stem, even when nonce affixes were used. However, reaction times during the decomposition of complex word forms with nonce affixes was modulated by the set size of the relevant affix categories. Indeed, set size is a statistical measure that is closely related to the transition probability between two categories, e.g. stems and affixes. Reaction times were also affected by vowel harmony. Non-harmonizing, but existent affixes evoked longer reaction times in a word recognition task than did nonce affixes, which the authors interpreted as evidence for a triple mechanism,

[3] This can be modified by different cliticization and other morphological phenomena (for discussions, see Kabak & Vogel, 2001; Peperkamp & Dupoux, 2002), but the observation remains true in most cases.

which first decomposes the complex form, then makes lexical decisions about the individual morphemes, and finally reassembles them to check for correct harmony and allomorphy.

In general, the idea that statistical segmentation might slightly overgenerate word internal morpheme boundaries, thus favoring morphological decomposition, is consistent with the fast mapping (Carey, 1978) or 'less is more' (Newport, 1990; Endress & Bonatti, 2006) 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) shows 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. Statistical segmentation that posits a high number of boundaries favors such an analytic learning strategy.

Frequency

Frequency is one of the most commonly used measures in computational linguistics. Its effects on language processing, lexical retrieval and other psycholinguistic processes are also well-known (e.g. Segui et al., 1982; Alario, Costa, & Caramazza, 2002; Grainger, 1990). The experiments reported in the previous chapters have investigated how frequency contributes to language acquisition. Two domains were explored: segmentation and the acquisition of functors and word order.

It has been found that the most frequent syllables and bisyllables constitute lexical items, and using their boundaries, other lexical items can also be segmented out. This segmentation is highly accurate, but somewhat less complete than segmentation based on conditional probabilities.

The most important feature of frequency-based segmentation is that it builds a large vocabulary of function morphemes, free and bound. While infants do not use functors in their early productions, it has been shown that they recognize the most frequent ones—first as a generic category irrespective of the representational details of individual functors, then also representing them in full phonemic detail (Gerken et al., 1990; Shi, Cutler, et al., 2006; Shi, Werker, & Cutler, 2006; Shi et al., 1999; Shi & Werker, 2001). It has been proposed (e.g. Shi, Cutler, et al., 2006; Shi, Werker, & Cutler, 2006; Christophe, Millotte, Bernal, & Lidz, in press) that one possible role of these early functors is to facilitate the segmentation of content words. Since nouns are typically accompanied by determiners, case markers etc., verbs are often preceded or followed by auxiliaries, person, number or gender agreement markers, functors delimit their concomitant content words, and serve as indicators of their lexical category, e.g. if a word is preceded by the article *the*, then it is a noun.

In addition to signaling the boundaries and the lexical category of content words, functors also serve to indicate word order. The relative order of functors and content words correlates with the relative order of a number of other constituents in a language, such as the verb and its object, the complementizer and the embedded clause it introduces etc. Utterance boundaries are especially informative in this regard, since they are universally available perceptual boundaries, where the order of functors and content words can be directly observed even by very young infants. In the experiments of Chapter 3, it has been shown that the relative order of functors and content words at utterance boundaries reliably correlates with the general word order properties of two typologically different languages, Japanese and Italian.

The proposal that frequent functors cue word order in early language acquisition is consistent with previous findings in artificial language learning studies with adults and children, showing that the presence of frequent marker elements significantly contributes to the learnability of positional regularities (the anchoring hypothesis, Braine, 1963, 1966; Green, 1979; Valian & Coulson, 1988; Valian & Levitt, 1996).

To more directly test the hypothesis that frequent functors play a role in the acquisition of word order, in Experiments 7 and 8 of Chapter 4, I tested adults' and prelexical infants' word order preferences in an artificial grammar learning paradigm. The artificial grammar was devised in such a way that the familiarization stream, being structurally ambiguous, could be assigned two opposite word orders: a [frequent item – infrequent item] vs. an [infrequent item – frequent item] order. As predicted by the hypothesis, adults speakers of five, typologically different languages, Basque, Japanese, Hungarian, Italian and French, show a word order preference that is consistent with the dominant word order of their native language in the two-alternative forced choice test following familiarization. However, adults have had a long experience with their native language and have a full blown vocabulary. This is not the case with 7-month-old infants, who are still in the prelexical stage of linguistic development. Yet, 7-month-old Japanese and Italian infants also show word order preferences consistent with the order of their future native language. Even more pronouncedly than adults, Japanese infants opt for an infrequent–frequent order, while Italian babies prefer frequent-infrequent patterns. These findings provide evidence that infants have at least a rudimentary representation of word order before they know their lexicon. This argues against theories (Tomasello, 2000) claiming that the acquisition of syntactic structure is dependent on the acquisition of individual lexical items, from which first semi-abstract patterns and later rules are extracted by analogical processes. Clearly, if word order is represented prior to lexical items, such a piecemeal, item-based learning is not possible.

A question that remains to be answered is how this frequency-based early representation of word order relates to more abstract word order rules, e.g. the Head-Complement parameter. While I cannot provide a definitive solution here, one possibility is that this representation conspires with other, perceptual cues to word order, e.g. the place and realization of prosodic prominence within phonological phrases (the Rythmic Activation Principle, Nespor et al., 1996, under review), in order to trigger the relevant setting of the word order parameters.

5.1.3 Learning the rules of natural language

One of the greatest challenges in language acquisition is to explain how abstract, complex and highly domain specific rules such as the ones that characterize adult grammars develop in the infant mind. The experiments in Chapter 2 have attempted to explore the beginnings of this rule learning process.

The results provide evidence that the newborn brain shows a signature for extracting a generalization from linguistic input when its structure conforms to a perceptual Gestalt. This is not the case when structurally similar, but perceptually not salient patterns are used. Thus, the ability to generalize a pattern is present in the initial state, although with a limited scope.

This finding raises a series of questions that future research will have to address. First, the nature of the limitation: do non-adjacent repetitions not function as perceptual Gestalts because they require more complex, more abstract identity computations or because the detection of identity is initially limited to neighboring stimuli?

Second, are generalizations based on perceptually salient patterns limited to linguistic stimuli or they are present in other modalities, too? Studies with older infants suggest that they can learn repetition sequences in vision (Saffran et al., in press), while repetitions in non-linguistic auditory stimuli are learned only when infants received prior training with linguistic sequences (Marcus et al., in press). To test whether newborns are able to detect repetitions in non-linguistic auditory sequences, future experiments will be designed that compare learning repetition-based linguistic sequences with learning similarly structured tone sequences.

A third question concerns the stability of the learned regularities. Establishing stable, long term representations is a necessary prerequisite for any learning process that is to play a role in language acquisition. Can a repetition-based regularity be retained in memory over a longer period of time? This issue may be explored by retesting infants a day after the initial exposure, using familiar and novel items both in the repetition and in the control condition. Such a paradigm would allow to investigate the memory traces of individual items, as well as generalizations.

A fourth, and more challenging, question is how rule extraction based on perceptual Gestalts relates to abstract, symbolic rules. What is the developmental trajectory of the ability to generalize? Does perceptually based rule learning represent a generalization mechanism that is different from symbolic rule learning or is it a developmental precursor to it? The paradigm used in the experiments of Chapter 2 offers a way to explore these issues by comparing repetition-based and more complex rules in newborns and older infants.

The remaining questions notwithstanding, evidence has been found indicating that the perceptual and computational system of the brain is ready to discover certain types of regularities in the surrounding linguistic environment.

5.2 The interactions between statistics, perception and rules

The field of language acquisition was for a long time divided between two extreme positions. The empiricist and constructivist views of connectionists (Elman et al., 1996) and some developmentalists (Tomasello, 2000) opposed to the rationalist stance of linguists (Chomsky, 2000, 2004, 1959). Recently, however, a synthesis has started to emerge, emphasizing the importance of both input-driven and rule-guided mechanisms (e.g. Golinkoff & Hirsh-Pasek, 2007). In a similar spirit, this section will highlight some of the interactions between statistical learning, perceptual Gestalts and rule learning observed in the experiments of the previous chapters, arguing for a view of language acquisition according to which different computational components of the human mind interact with and provide input for the language acquisition faculty.

In Chapter 2, it has been observed that the newborn brain can draw generalizations from stimuli honoring configurations that can be efficiently encoded by the perceptual system. A different regularity, which is equal in symbolic complexity, but does not represent a perceptual Gestalt, cannot be learned.

Rule learning has also been found to be limited by other perceptual aspects of the signal (Peña et al., 2002). When a speech stream is perceived as continuous, it triggers statistically based segmentation, but does not allow generalizations to be made. Only when the input is already segmented can structural regularities be extracted from the segmented chucks. This finding is highly relevant for the acquisition of morphological regularities in agglutinating languages, since it argues that in order for complex morphological regularities to be learnable, the input has to be perceived as already segmented. Of course, agglutinating languages do not have pauses after every word, not even after polymorphemic ones. But, as described above, these languages typically have several morpho-phonological processes that signal word boundaries. One possible role of these processes during language acquisition might be to perceptually segment the input, thus triggering generalizations instead of statistical segmentation.

It is interesting to note in this respect that perceptual Gestalt principles also limit the scope of statistical learning. In a series of studies, Aslin, Newport and colleagues have shown that statistically based segmentation both in the auditory and in the visual domain obey the principles of similarity (Creel, L., & Aslin, 2004), proximity (Peña et al., 2002; Newport & Aslin, 2004) or good continuation (Fiser, Scholl, & Aslin, 2007).

Another area where the experiments of the thesis have revealed interactions between statistics, perceptual cues and generalizations is the acquisition of functors and the related problem of learning word order. As discussed before, some rudimentary representation of word order is established in prelexical infants on the basis of the relative order of frequent and infrequent words in the language of exposure. I have hypothesized that this representation is convergent with another one, established on the basis of the position and acoustic realization of prosodic prominence in phonological phrases (Nespor et al., under review). Carrying out acoustic measurements in a number of languages, the authors show that in Object-Verb languages, the prominence is at the left edge of the phonological phrase and it is realized by an increase in pitch and intensity, whereas in Verb-Object languages, the prominence is right-most and is implemented as lengthening. Thus, in an OV language, the speech stream is an alternation of higher and lower intensity elements, while in VO languages, what alternate are shorter and lengthened units. Given the iambic-trochaic principle of auditory grouping (B. Hayes, 1995), according to which elements contrasting in intensity naturally group with initial prominence, while elements contrasting in duration group with final prominence, young learners have a perceptually available mechanism to bracket the signal into constituent phrases. Since the position of prosodic prominence correlates with word order, this prosodic bracketing is a possible bootstrapping cue to word order. Nespor et al. (1996) have shown that infants are able to use these prosodic differences to distinguish French, a VO language from Turkish, an OV language, even when other phonological information has been removed (e.g. phonemes etc.). Since both the frequency based and the prosodically based triggering mechanisms lead to phrase bracketing of a very similar type, it is not unreasonable to believe that they converge towards or feed into the same abstract word order representation.

Looking at the above findings from a temporal perspective, it is clear that the interactions between the mechanisms start immediately after birth and continue through the prelinguistic development of the infants (and presumably, even further). This is to be expected, since as many authors have emphasized (Morgan & Demuth, 1996), cue in the signal are most efficient when considered in conjunction (the theory of "convergent cues"). Consequently, the different mechanisms that process these cues need to interface throughout linguistic development.

The above findings provide evidence for complex interactions between the different learning mechanisms investigated in this thesis. From a logical point of view, these interactions might be understood as a necessary consequence of language being a mediator between sound/sign and meaning, or, from a developmental perspective, of the linking problem in language acquisition (Pinker, 1984).

On the one hand, the language acquisition device receives its input from the environment indirectly, through other cognitive components, such as perception—the auditory system in the case of speech and vision in the case of sign language. In other words, perception creates a link between the input and the language faculty. However, this link is far from being a straightforward one-to-one mapping. Perceptual mechanisms process the input according to their own biases and encode it in their own representations—operations that have often been ignored in language acquisition. What serves as raw material to the language faculty is this perceptually 'filtered' input. This view of the interaction between perception and the language faculty is somewhat reminiscent of early proposals about parsibility (Kimball, 1973) as a condition of linguistic representations.

On the other hand, the representations and computations of the language faculty cannot simply be reduced to such perceptual biases. For instance, bracketing the signal on the basis of prosodic prominence or the position of frequent items does not directly predict that in English the word *that* needs to precede the phrase *you moved to Canada* when producing or comprehending the sentence *I heard that you move to Canada.*. For this knowledge to be available to the langauge learner, categories such as complementizer, subordinate clause and rules such as the Head-Complement parameter or equivalent are necessary. Such representations may seem superfluous, not parsimonious enough or violating Ockham's razor. But only if we ignore the fact that language interfaces not only with perception, but also with the conceptual system—semantics, world knowledge, the language of though etc. This system has its own representational code, and thus places constraints on the input that the language faculty needs to provide for it. While the mapping between language and the conceptual system is less well understood than between language and the perceptual and production systems, it is reasonable to believe that the former, just as the latter, has its own representational requirements.

At the beginning of the thesis, I distinguished two approaches that address the question of why language is unique to humans and how it integrates into the human mind. The approach I have chosen was to explore the ontogenesis of language in early development. The conclusions arrived at are not, however, alien to the evolutionary approach, either. A view of language that emphasizes the interactions between different components of the mind and shows how perceptual, general and specific computations conspire to give rise to human language bears close resemblance to biological and evolutionary theories of animal and human cognition that argue for the presence of specific mechanisms, dedicated to the heuristic solution of given problems, acting in concert (Gallistel, 1990, 2000). Models of language evolution (Hauser, Chomsky, & Fitch, 2002 and related debates in Fitch et al., 2005; Pinker & Jackendoff, 2005) claiming that language recruited some already existing perceptual and general abilities, connecting them with representations and computations evolved to specifically subserve language are also compatible with the results obtained here. This convergence between the two approaches is an expected and welcome result, suggesting that interdisciplinary investigations of both kinds should proceed hand in hand.

5.3 Summary of main findings

The experiments of the thesis have provided empirical and theoretical support for the four main hypotheses proposed in the Introduction (Chapter 1). Therefore, I conclude that:

1. Humans are born equipped with auditory computational primitives that allow them to process and learn certain structural aspects of auditory stimuli immediately at birth. Such primitives can pave the way for other perceptual and symbolic computations that play a role later during language acquisition. [Experiments 1–3, Chapter 2]

2. The input that young learners receive is rich in—statistical, prosodic etc.—information that correlates with and potentially bootstraps structural categories and regularities. [Experiments 4–6, Chapter 3]

3. Infants are able to use this information to learn about structure, e.g. word order, independently of and prior to the development of the lexicon. [Experiments 7–8, Chapter 4]

4. During language acquisition, the genetically endowed abstract linguistic knowledge develops to match the target grammar by relying on information contained in the input and representations sanctioned by the perceptual system (in addition to other, mostly maturational processes).

Chapter 6

Conclusion

This thesis has explored certain properties of the linguistic input and some abilities of language learning infants in an attempt to deepen our understanding of language acquisition. In particular, it has investigated perceptual Gestalts, statistical learning and rule learning, and it has shown how particular aspects of language serve as specific input to these mechanisms. The resulting picture is that of a speech signal rich in information, which provide input for a range of different learning mechanisms, some not specific to language.

At a first glance, this is quite different from the description of language acquisition as an inductive impossibility, which served as the starting point of my investigations. The cornerstone of that view is the 'poverty of stimulus' argument, claiming that the input does not contain information about the underlying rule system that generates it. The results obtained here, in contrast, seem to suggest that the signal is rich in cues. Were the initial assumptions wrong, then?

Not necessarily. The contradiction between the logical framework defined by the poverty of stimulus argument and the obtained results is only apparent. The poverty of stimulus argument states that there is no explicit information about *structure*, or more specifically about the rules that generate the linguistic input. What has been found in this thesis, in accordance with other works in the bootstrapping framework, is that the input contains *non-structural* information that correlates with some structural properties. However, without the knowledge of the correlations, the cues remain uninformative. Consequently, the input *is* poor in structural information. Even more importantly, a signal that is rich in statistical, prosodic, phonological etc. information, correlated with structure, requires a priori knowledge of the correlations in order for the cues to be useful indicators of structure. Therefore, the enriched signal does not render a priori knowledge unnecessary. On the contrary, it can only be processed by dedicated mechanisms that have the relevant cue–structure correspondences encoded.

What is the nature of these mechanisms? Some are specific to the linguistic domain. The syntactic bootstrapping strategy, which deduces the syntactic/semantic type of a verb from the number and order of the noun phrases it co-occurs with, has no application outside language.

Other mechanisms apply to linguistic and non-linguistic stimuli alike. Statistical computations are readily performed in the auditory and visual domains. The iambic-trochaic grouping principle applies to prosodic prominence, as well as ambulance sirens. Repetitions and edges are salient both in audition and vision. Yet, even these more general, often perceptually based mechanisms need to map onto language-specific representations in order to bootstrap grammar. Statistics cannot be computed without language-specific combinatorial units such as syllables, consonants, phonetic features, over which to compute them. Prosodic patterns cannot be grouped into iambs or trochees without the notion of a domain in which the grouping applies, i.e. the phonological phrase. Conditional probabilities segment out units within which morphological regularities can be discovered. Edges appear because utterances can be broken down into prosodic and syntactic constituents.

What emerges, then, is a view of language acquisition as a process of decryption that decodes a myriad of cues of various nature in the linguistic signal by integrating the outputs of statistical, perceptual and linguistic mechanisms with a priori, language-specific representations to arrive at the underlying rules that generate the signal.

‘

Appendix A

Appendix to Chapter 3

The appendix discusses some issues that were not included in the main text of Chapter 3, because they are not directly relevant for language acquisition, yet they deserve some attention from a computational point of view. Most of these questions relate to the frequency distribution of words, syllables and conditional probability values in large language corpora, and specifically to Zipf's law.

A.1 Zipf's law

Word frequency is the statistical measure whose distribution has received the most attention in computational linguistics (Cancho, 2005a, 2005b; Cancho & Sole, 2001; Zipf, 1935). According to Zipf's original observation, known as Zipf's law, word frequencies follow a distribution that can be described by a power law. Somewhat formally,

$$F_n \sim n^{-a}, \tag{A.1}$$

where F_n is the frequency of the nth element in the frequency hierarchy, i.e. the rank of the given element, and a is close to 1. Intuitively, this means that a few elements occur very frequently, while the majority of them occur much more rarely; and that the frequency of an element is roughly inversely proportional to its rank. Plotting the logarithm of the frequency against the logarithm of the rank, an approximately linear relation obtains (Figure A.1, adapted from (Cancho & Sole, 2003)).

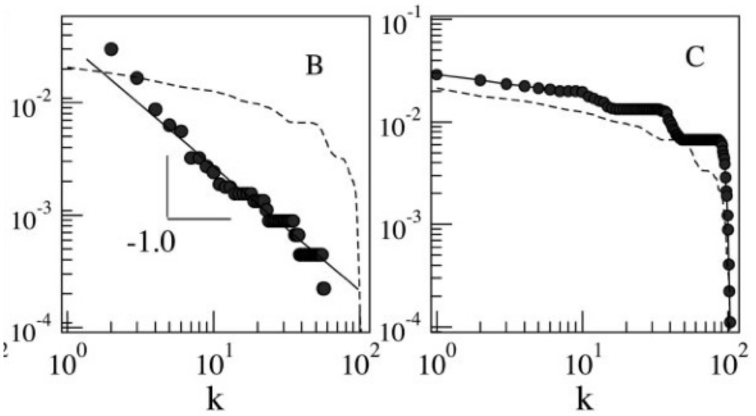


Figure A.1: Two Zipfian distributions: the linear relation between the logarithm of the frequency and the logarithm of the rank of a word in the frequency list. B: The ideal case of human language. C: Two regimes in the Zipfian distribution of word frequencies.

This distribution is commonly found in many different natural phenomena from population sizes of cities to file sizes on the internet. Since Zipf's original observation, it has been repeatedly found that word frequencies follow this distribution in many different languages (Cancho, 2005a, 2005b; Cancho & Sole, 2001). Such a distribution has two consequences for the investigations pursued in Chapter 3.

First, it is not a distribution that naturally presents several modes. The only one is the frequency of the first, most frequent element. Second, this distribution demonstrates that sparsity is an inherent, inevitable property of large corpora. Some items are very frequent, while others are rare. The actual sparsity depends on a number of factors (e.g. the size of the corpus), but some items will necessarily be infrequent.

Depending on the extent of sparsity, small deviations from the linear relation can sometimes be observed, almost as if two linears, rather than one could be fit on the curve (Figure A.1). In fact, it has been claimed (Cancho & Sole, 2001, 2003; Cancho, 2005a, 2005b) that in the case of the human lexicon, the split between the two linears or 'regimes' is a meaningful distinction. It is argued to reflect a communicative phase transition between a small vocabulary of very frequent items shared by all speakers of the same language community and a larger vocabulary of infrequent and more specialized terms, varying from one speaker to the other as a function of occupation, experience, social class etc.

A.2 Hungarian and Italian TPs distribute according to Zipf's law

Given the ubiquity of Zipfian distributions in natural phenomena, it is not surprising that TP[1] values also distribute this way in large language corpora. Here, I will demonstrate that the TP distributions in the Hungarian, small Italian and full Italian corpora do indeed follow Zipf's law. Then, I will discuss some of the consequences of this observation.

In order to demonstrate that the TP distributions derived from the three infant-directed corpora are Zipfian, it needs to be shown that a linear relation holds between the logarithm of the frequency of words and the logarithm of their rank in the frequency hierarchy. Figures A.2–A.4 illustrate that the linear relation between the logarithms of frequency and rank holds in all three corpora. It can also be observed that, as discussed above, sparsity causes deviations from linearity, creating 'two regimes' in the sparser Hungarian corpus. Whether these two regimes have some communicative function, as in the case of word frequencies, is an open question.

A.2.1 Sparsity

Since sparsity causes deviations from the perfect Zipfian distributions, it is important to evaluate its effects in some detail and show that the deviations found in the corpora are indeed due to sparsity. In order to do this, I have recomputed the TP and MI distributions derived from the Hungarian corpus,

[1] For abbreviations, see p. 1.

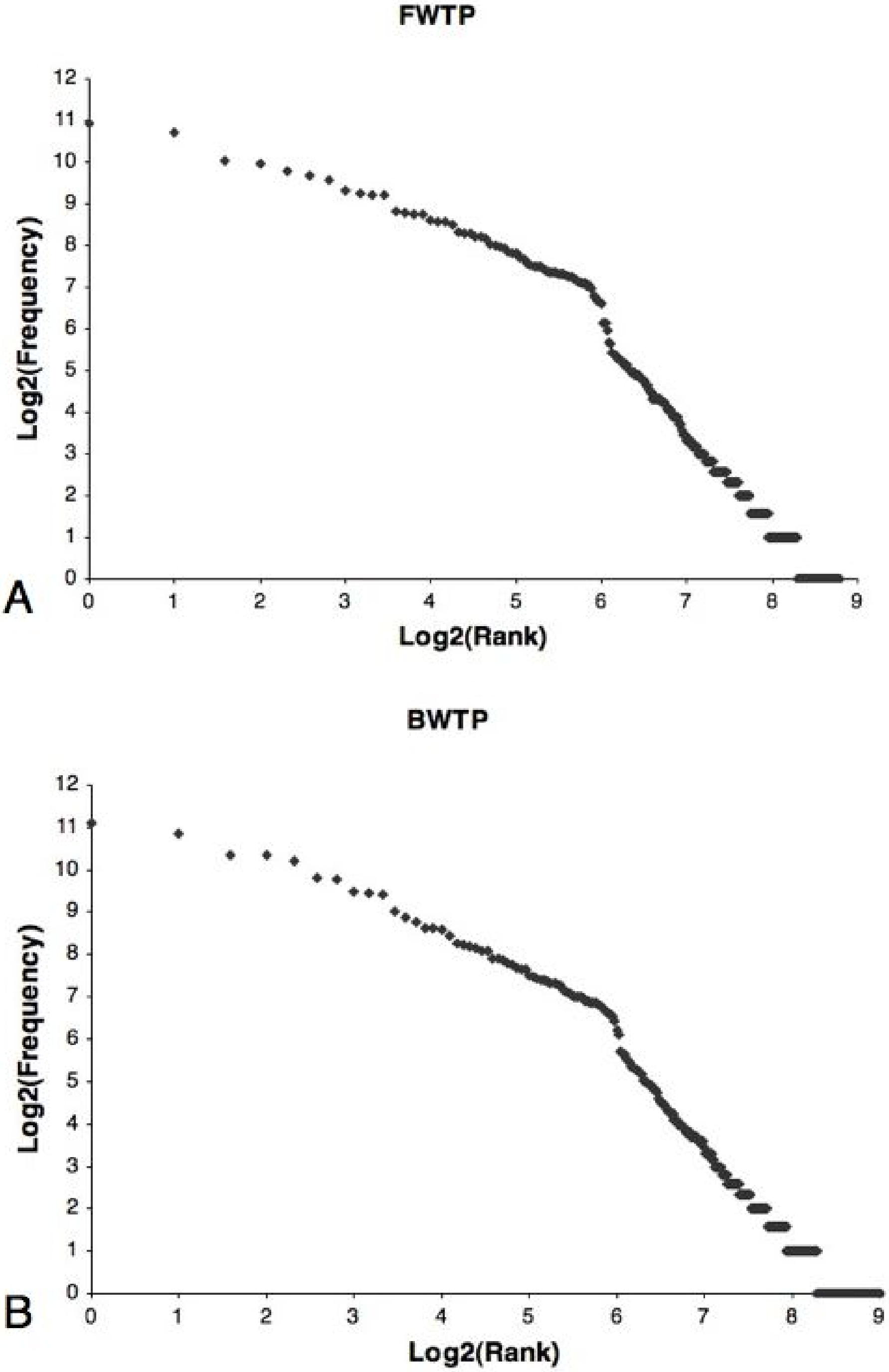


Figure A.2: The Zipfian distribution of TP values in the Hungarian corpus. A: FWTPs. B: BWTPs.

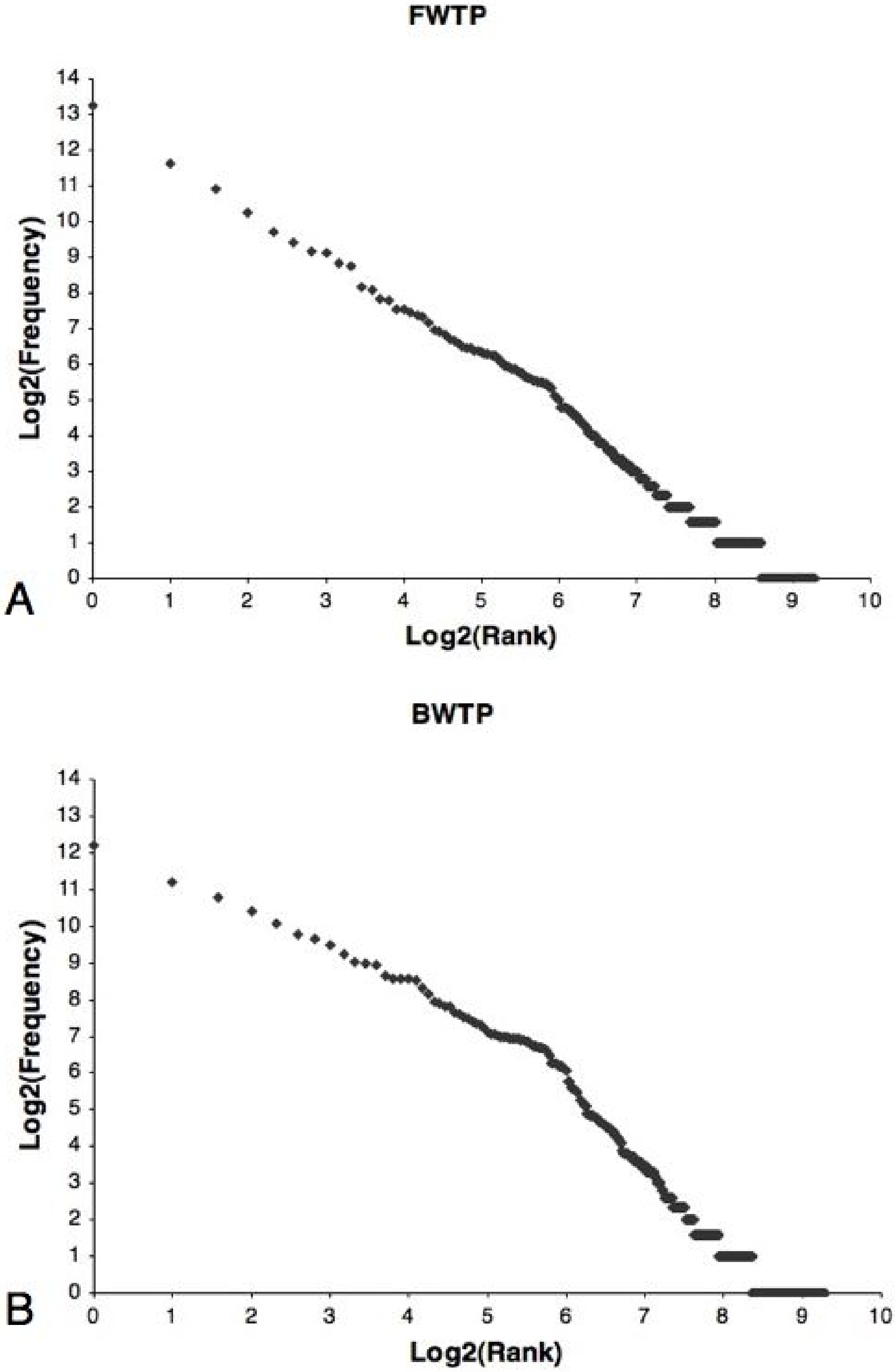


Figure A.3: The Zipfian distribution of TP values in the full Italian corpus. A: FWTPs. B: BWTPs.

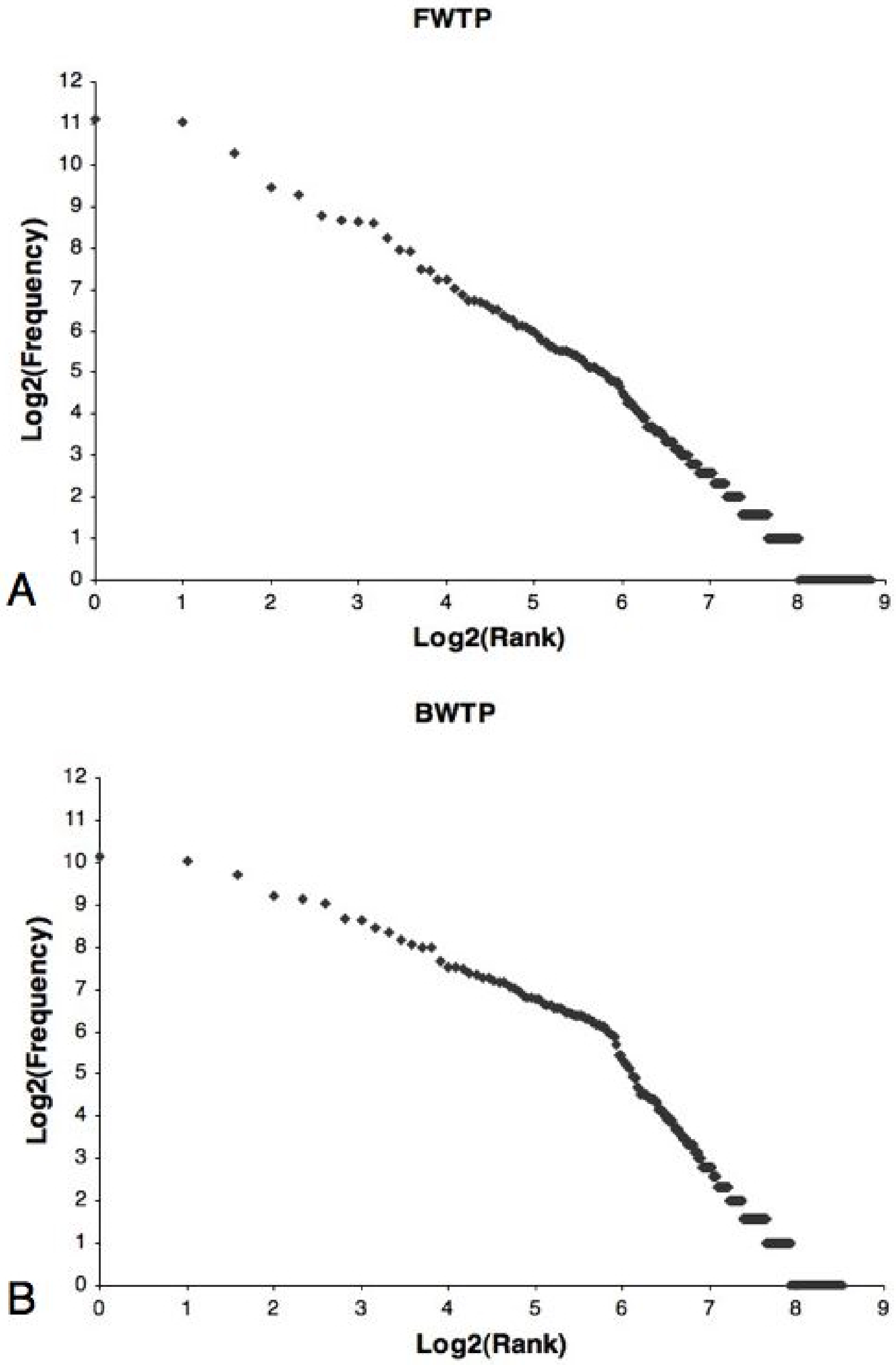


Figure A.4: The Zipfian distribution of TP values in the small Italian corpus. A: FWTPs. B: BWTPs.

which is the sparsest of the three, using a sparsity threshold of 10[2]. TP and MI values obtained from syllable pairs in which at least one member appears with a frequency of less than 10 was considered a sparse datum. As Figures A.5 and A.6 illustrate, when plotted in separate distributions, it is clearly observable that the 'deviant' peaks are due to sparse data.

A.2.2 Percentiles defining the set of arbitrary thresholds

Zipfian distributions are not multimodal. No threshold values could be derived from them for segmentation, so a set of arbitrary thresholds were obtained as the 1st, 2nd, 3rd etc. percentiles of the FWTP, BWTP and MI distributions. The following tables indicate the exact values for each distribution in the two languages.

[2] This value was chosen arbitrarily. Other values, e.g. 5, 100 and even 270, were also tested and yielded very similar results, although the amount of sparse data obviously increased by increasing the threshold.

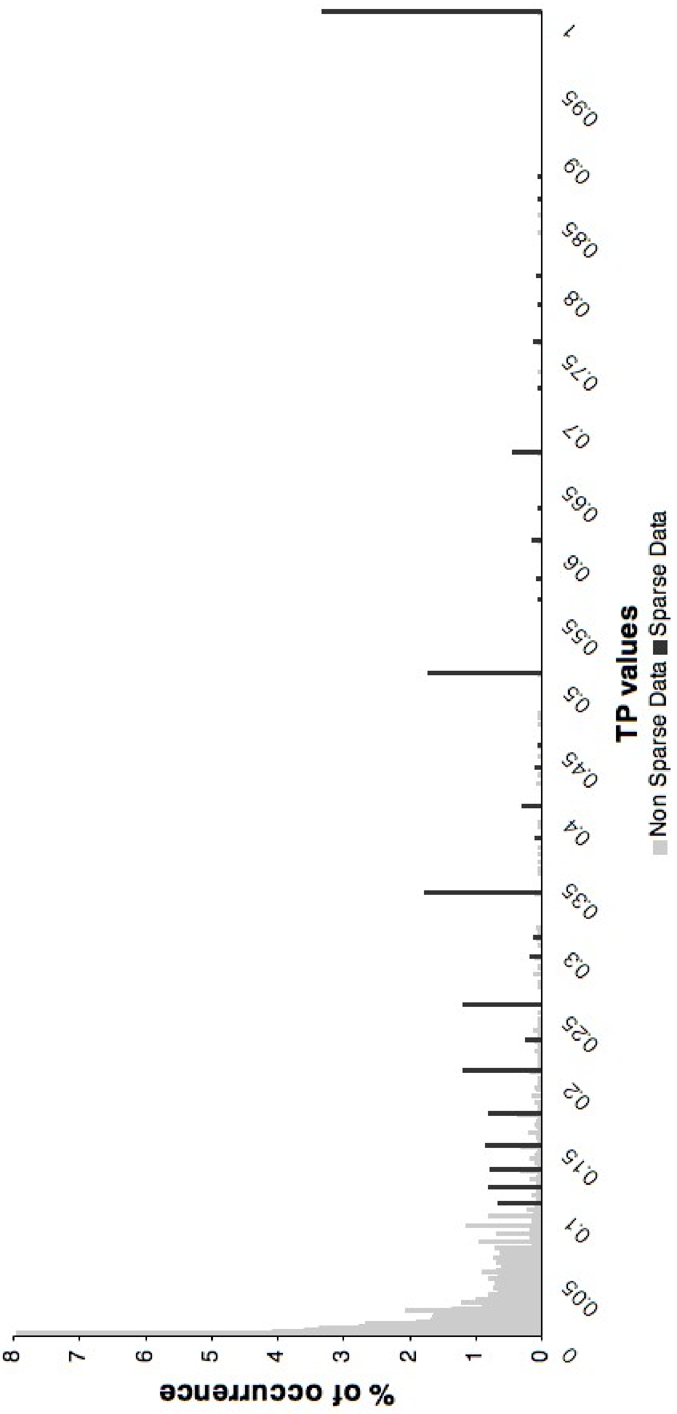


Figure A.5: Sparse and non-sparse data in the Hungarian FWTP distribution. Light grey bars indicate non-sparse data, dark grey bars indicate sparse data.

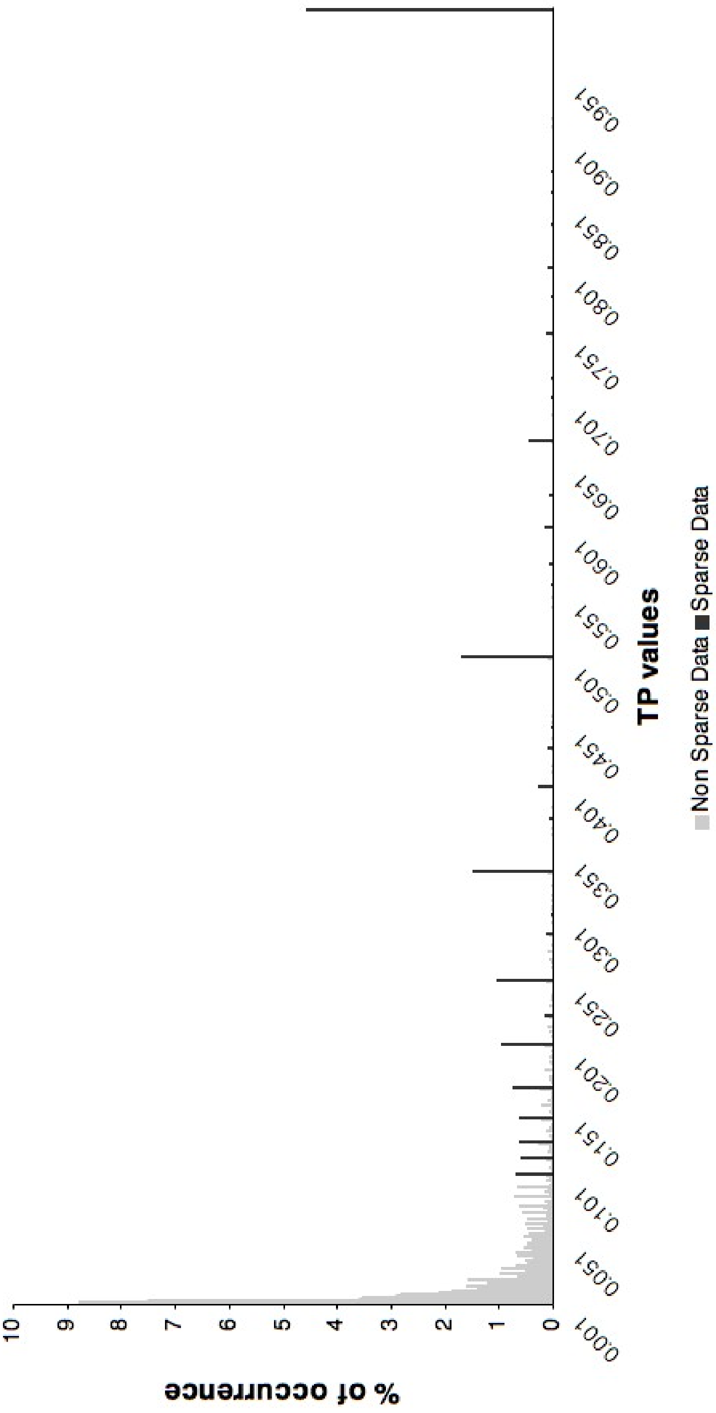


Figure A.6: Sparse and non-sparse data in the Hungarian BWTP distribution. Light grey bars indicate non-sparse data, dark grey bars indicate sparse data.

Percentile	FWTP value
1	0.000382117
2	0.000573175
3	0.000850702
4	0.000982318
5	0.001058201
6	0.001146351
7	0.001295337
8	0.00137931
9	0.001545595
10	0.001683502
11	0.001788909
12	0.001984127
13	0.002116402
14	0.002173913
15	0.002403846
16	0.002518892
17	0.002695418
18	0.00286533
19	0.003006012
20	0.003236246
21	0.003484321
22	0.003703704
23	0.00400534
24	0.004255319
25	0.004608295
26	0.004878049
27	0.005050505
28	0.005340454
29	0.005813953
30	0.006134969
31	0.006535948
32	0.006877879
33	0.007326007
34	0.007751938
35	0.008196721
36	0.008547009
37	0.008979748
38	0.009259259
39	0.00990099
40	0.010204082
41	0.010869565
42	0.011594203
43	0.012048193
44	0.012604619
45	0.013333333
46	0.014234875
47	0.015151515
48	0.015748031
49	0.016574586
50	0.017241379

Table A.1: The FWTP values used as arbitrary thresholds in Hungarian.

Percentile	FWTP value
51	0.018050542
52	0.018518519
53	0.019545486
54	0.02020202
55	0.02173913
56	0.022727273
57	0.024096386
58	0.025423729
59	0.027027027
60	0.028571429
61	0.03030303
62	0.03125
63	0.033333333
64	0.035714286
65	0.0375
66	0.04
67	0.043380017
68	0.045454545
69	0.047619048
70	0.050505051
71	0.054545455
72	0.058823529
73	0.0625
74	0.066666667
75	0.071428571
76	0.076923077
77	0.083333333
78	0.089112447
79	0.095238095
80	0.1
81	0.111111111
82	0.125
83	0.134328358
84	0.142857143
85	0.166666667
86	0.172904509
87	0.2
88	0.217391304
89	0.25
90	0.272727273
91	0.333333333
92	0.333333333
93	0.4
94	0.5
95	0.5
96	0.666666667
97	1
98	1
99	1
100	1

Table A.2: The FWTP values used as arbitrary thresholds in Hungarian.

Percentile	BWTP value
1	0.000382117
2	0.000390016
3	0.000764234
4	0.000894454
5	0.000982318
6	0.00102145
7	0.001170047
8	0.001295337
9	0.00137931
10	0.001545595
11	0.001626016
12	0.001776199
13	0.001895735
14	0.002004008
15	0.002157497
16	0.002344666
17	0.002475248
18	0.00255102
19	0.002747253
20	0.00286533
21	0.002949853
22	0.003134796
23	0.003289474
24	0.003552398
25	0.003663004
26	0.003883495
27	0.0041841
28	0.004464286
29	0.004694836
30	0.004926108
31	0.005154639
32	0.005393743
33	0.005780347
34	0.006060606
35	0.006476684
36	0.006711409
37	0.007142857
38	0.007434944
39	0.0078125
40	0.008196721
41	0.008474576
42	0.008878682
43	0.009259259
44	0.00990099
45	0.010416667
46	0.010989011
47	0.011583333
48	0.012345679
49	0.013157895
50	0.013888889

Table A.3: The BWTP values used as arbitrary thresholds in Hungarian.

Percentile	BWTP value
51	0.014492754
52	0.015306122
53	0.016393443
54	0.017241379
55	0.018181818
56	0.018867925
57	0.02020202
58	0.02173913
59	0.023255814
60	0.024390244
61	0.025641026
62	0.027777778
63	0.029126214
64	0.030791072
65	0.033333333
66	0.035714286
67	0.037735849
68	0.04
69	0.043478261
70	0.047619048
71	0.050847458
72	0.055555556
73	0.059322034
74	0.066666667
75	0.071428571
76	0.076923077
77	0.083333333
78	0.090909091
79	0.1
80	0.111111111
81	0.125
82	0.13548328
83	0.146326719
84	0.166666667
85	0.187853774
86	0.2
87	0.25
88	0.25
89	0.333333333
90	0.333333333
91	0.410829563
92	0.5
93	0.5
94	0.666666667
95	0.875
96	1
97	1
98	1
99	1
100	1

Table A.4: The BWTP values used as arbitrary thresholds in Hungarian.

Percentile	MI value
1	-2.224082086
2	-1.665983461
3	-1.318184601
4	-1.037714349
5	-0.817832691
6	-0.620370944
7	-0.440644366
8	-0.276166082
9	-0.116883426
10	0.024290691
11	0.154725543
12	0.287073932
13	0.399457877
14	0.510665217
15	0.611619909
16	0.721753912
17	0.822430078
18	0.928617756
19	1.031930314
20	1.14301356
21	1.244064621
22	1.344658929
23	1.431643743
24	1.52933739
25	1.613465095
26	1.708449504
27	1.799148745
28	1.878310432
29	1.961526645
30	2.051512184
31	2.145249052
32	2.226032999
33	2.323163574
34	2.412107406
35	2.492871923
36	2.578456114
37	2.65205517
38	2.750913315
39	2.83794347
40	2.923062038
41	3.008140389
42	3.09606524
43	3.182707358
44	3.255363155
45	3.348603913
46	3.431806255
47	3.526627423
48	3.618219379
49	3.705312393
50	3.79528209

Table A.5: The MI values used as arbitrary thresholds in Hungarian.

Percentile	MI value
51	3.892029182
52	3.992287542
53	4.086183289
54	4.175886092
55	4.228498706
56	4.32036569
57	4.40603397
58	4.51458393
59	4.617964191
60	4.717940001
61	4.832121094
62	4.94472276
63	5.044513107
64	5.148919045
65	5.272702264
66	5.370243198
67	5.476366648
68	5.601107447
69	5.723897457
70	5.844797696
71	5.958948101
72	6.092335516
73	6.230732633
74	6.363613916
75	6.503952035
76	6.622850233
77	6.758053032
78	6.9074178
79	7.045825552
80	7.183528081
81	7.318844046
82	7.470847139
83	7.646795928
84	7.847223018
85	8.016281276
86	8.197861436
87	8.41837972
88	8.627847617
89	8.833417219
90	9.05580964
91	9.262260518
92	9.516298103
93	9.792775234
94	10.10981437
95	10.52529492
96	10.96270024
97	11.54766274
98	12.2491112
99	13.37773774
100	16.54766274

Table A.6: The MI values used as arbitrary thresholds in Hungarian.

Percentile	FWTP value
1	0.000361402
2	0.000382848
3	0.000422654
4	0.000439174
5	0.000543774
6	0.000587372
7	0.000626566
8	0.000665779
9	0.000721501
10	0.000754148
11	0.000822368
12	0.000856898
13	0.000884956
14	0.000892061
15	0.00094697
16	0.001004016
17	0.001044932
18	0.001119821
19	0.001176471
20	0.001231527
21	0.001293661
22	0.001372684
23	0.001443001
24	0.001497006
25	0.001540832
26	0.001631321
27	0.001672241
28	0.001713796
29	0.001792115
30	0.001893939
31	0.002008032
32	0.002164502
33	0.002239642
34	0.002364066
35	0.00243309
36	0.002512563
37	0.002635046
38	0.002770083
39	0.002904163
40	0.002997619
41	0.003231018
42	0.003328895
43	0.003568243
44	0.003768554
45	0.003966384
46	0.004147813
47	0.004424779
48	0.004576659
49	0.004846527
50	0.005089062

Table A.7: The FWTP values used as arbitrary thresholds in Italian.

Percentile	FWTP value
51	0.005491225
52	0.005882353
53	0.006244425
54	0.006525285
55	0.006798679
56	0.007228916
57	0.007407407
58	0.007789255
59	0.008237232
60	0.008823529
61	0.00913242
62	0.009812667
63	0.010500808
64	0.011235955
65	0.011764706
66	0.012313312
67	0.013392867
68	0.014681892
69	0.015625
70	0.01676353
71	0.018181818
72	0.019417476
73	0.020833333
74	0.022613065
75	0.024390244
76	0.027027027
77	0.030518021
78	0.033333333
79	0.0375
80	0.042105263
81	0.047619048
82	0.05435012
83	0.064516129
84	0.073170732
85	0.086956522
86	0.1
87	0.113570878
88	0.142857143
89	0.166666667
90	0.2
91	0.25
92	0.3
93	0.348742432
94	0.479744715
95	0.516185785
96	0.668291032
97	0.957927327
98	1
99	1
100	1

Table A.8: The FWTP values used as arbitrary thresholds in Italian.

Percentile	BWTP value
1	0.000382848
2	0.000531915
3	0.000626566
4	0.000721501
5	0.000807754
6	0.000892061
7	0.00094697
8	0.001044932
9	0.001174743
10	0.001328021
11	0.001451379
12	0.001540832
13	0.001647446
14	0.001762115
15	0.001881468
16	0.002089864
17	0.002283105
18	0.002409639
19	0.002512563
20	0.002676182
21	0.002886003
22	0.003062787
23	0.003294893
24	0.003448276
25	0.003723404
26	0.004016064
27	0.004366812
28	0.004552352
29	0.004734848
30	0.004986406
31	0.005309735
32	0.005801823
33	0.006024096
34	0.006329114
35	0.006521188
36	0.006788559
37	0.007136485
38	0.007371007
39	0.007751938
40	0.008232972
41	0.00872093
42	0.009049774
43	0.009456265
44	0.010102515
45	0.010638298
46	0.011160059
47	0.011627907
48	0.012110726
49	0.012787724
50	0.013333333

Table A.9: The BWTP values used as arbitrary thresholds in Italian.

Percentile	BWTP value
51	0.014084507
52	0.015075377
53	0.015686275
54	0.016483516
55	0.01758794
56	0.01826484
57	0.019230769
58	0.02
59	0.020833333
60	0.02173913
61	0.02247191
62	0.023668639
63	0.025
64	0.026030369
65	0.027303754
66	0.029427078
67	0.031121684
68	0.032258065
69	0.033996676
70	0.036231884
71	0.038464611
72	0.041666667
73	0.043478261
74	0.046511628
75	0.05
76	0.052631579
77	0.055555556
78	0.060606061
79	0.064516129
80	0.069726938
81	0.075397743
82	0.083333333
83	0.090209642
84	0.094707419
85	0.102040816
86	0.111111111
87	0.125
88	0.139250646
89	0.15
90	0.166666667
91	0.2
92	0.25
93	0.285714286
94	0.333333333
95	0.428571429
96	0.5
97	0.666666667
98	1
99	1
100	1

Table A.10: The BWTP values used as arbitrary thresholds in Italian.

Percentile	MI value
1	-3.382135067
2	-2.887419102
3	-2.552026874
4	-2.301158162
5	-2.105860362
6	-1.926364473
7	-1.770162565
8	-1.626781301
9	-1.517179305
10	-1.403358158
11	-1.299434717
12	-1.200884773
13	-1.107796735
14	-1.001647782
15	-0.912676118
16	-0.82278242
17	-0.738379243
18	-0.661264299
19	-0.597108427
20	-0.512343926
21	-0.449737844
22	-0.388839682
23	-0.317103205
24	-0.252049101
25	-0.181481254
26	-0.114621619
27	-0.052184627
28	0.016505303
29	0.08534824
30	0.151420822
31	0.214452521
32	0.274659052
33	0.347346256
34	0.409617773
35	0.470566286
36	0.536417036
37	0.607490863
38	0.667346196
39	0.742161008
40	0.80139505
41	0.865555509
42	0.936539859
43	0.999941026
44	1.063974478
45	1.137308939
46	1.214468546
47	1.285720329
48	1.357781735
49	1.434145603
50	1.507452517

Table A.11: The MI values used as arbitrary thresholds in Italian.

Percentile	MI value
51	1.57189161
52	1.653519429
53	1.73665176
54	1.822892217
55	1.895400862
56	1.97773118
57	2.073335064
58	2.152299555
59	2.24438777
60	2.329165236
61	2.41933092
62	2.498144156
63	2.584884036
64	2.662524103
65	2.76011502
66	2.850115833
67	2.97489062
68	3.070395605
69	3.185282466
70	3.289803079
71	3.421190469
72	3.538786141
73	3.665528194
74	3.769123167
75	3.892524426
76	4.03751303
77	4.148659744
78	4.274363224
79	4.392630511
80	4.52740495
81	4.695051261
82	4.82780647
83	5.022310055
84	5.137308939
85	5.280013762
86	5.447028838
87	5.646062926
88	5.822912203
89	6.040902426
90	6.266660999
91	6.443854001
92	6.658766351
93	6.998313454
94	7.321455599
95	7.830036638
96	8.347572258
97	9.095930698
98	9.734624293
99	11.15538766
100	16.48824812

Table A.12: The MI values used as arbitrary thresholds in Italian.

References

Abney, S. (1987). *The English noun phrase in its sentential aspect.* Unpublished doctoral dissertation, MIT.

Alario, F. X., Costa, A., & Caramazza, A. (2002). Frequency effects in noun phrase production: Implications for models of lexical access. *Language and Cognitive Processes*, *17*(3), 299–319.

Ambridge, B., Rowland, C. F., & Pine, J. M. (in preparation). Is structure dependence an innate constraint?

Antal, L. (1977). *Egy új magyar nyelvtan felé.* Budapest: Magvető.

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.

Austin, A., Newport, E. L., & Wonnacott, E. (n.d.). Predictable versus unpredictable variation: Regularization in adult and child learners. *BU-CLD 31, 3–5 November 2006, Boston, MA, USA.*

Baker, M. (2001). *The atoms of language.* New York, NY: Basic Books.

Batchelder, E. O. (2002). Bootstrapping the lexicon: a computational model of infant speech segmentation. *Cognition*, *83*, 167–206.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representation of speech sounds. *Journal of Experimental Psychology: General*, *117*(1), 21–33.

Bickel, B., & Nichols, J. (2007). Inflectional morphology. In T. Shopen (Ed.), *Language typology and syntactic description.* Cambridge: Cambridge University Press.

Bickerton, D. (1990). *Language and species.* Chicago, IL: University of Chicago Press.

Bloomfield, L. (1933). *Language.* New York, NY: Holt.

Bloomfield, L. (1939). Menomini Morphophonemics. *Travaux du Cercle Linguistique de Prague*, *8*, 105–115.

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

Bradley, D. (1978). *Computational distinctions of vocabulary type.* Unpublished doctoral dissertation, Monash University.

Braine, M. D. (1963). On learning the grammatical order of words. *PsychologR*, *70*(4), 323–348.

Braine, M. D. (1966). Learning the positions of words relative to a marker element. *Journal of Experimental Psychology*, *72*(4), 532–540.

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

Brown, C., Hagoort, P., & Keurs, M. ter. (1999). Electrophysiological signatures of visual lexical processing: open- and closed-class words. *Journal of Cognitive Neuroscience*, *11*(3), 261–281.

Brown, R. (1973). *A First Language. The Early Stages.* Cambridge, MA: Harvard University Press.

Caesar, K., Gold, L., & Lauritzen, M. (2003). Context sensitivity of activity-dependent increases in cerebral blood flow. *Proc Natl Acad Sci U S A*, *100*, 4239–4244.

Caesar, K., Thomsen, K., & Lauritzen, M. (2003). Dissociation of spikes, synatic activity and activity-dependent incerements in rat cerebellar blood flow by tonic synaptic inhibition. *Proc Natl Acad Sci U S A*, *100*(26), 16000–16005.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *CPsy*, *33*, 111–153.

Cancho, R. Ferrer-i. (2005a). The variation of zipf's law in human language.

European Physical Journal B, *44*(2), 249–257.

Cancho, R. Ferrer-i. (2005b). Zipf's law from a communicative phase transition. *European Physical Journal B*, *47*(3), 449–457.

Cancho, R. Ferrer-i, & Sole, R. V. (2001). Two regimes in the frequency of words and the origins of complex lexicons: Zipf's law revisited. *Journal of Quantitative Linguistics*, *8*(3), 165–173.

Cancho, R. Ferrer-i, & Sole, R. V. (2003). Least effort and the origins of scaling in human language. *Proc Natl Acad Sci USA*, *100*(3), 788–791.

Caramazza, A., Miceli, G., Silveri, M., & Laudanna, A. (1985). Reading mechanisms and the organization of the lexicon: Evidence from acquired dyslexia. *CogN*, *2*, 81–114.

Caramazza, A., & Zurif, E. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *B&L*, *3*, 572–582.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality.* Cambridge, MA: MIT Press.

Carral, V., Huotilainen, M., Ruusuvirta, T., Fellman, V., Naatanen, R., & Escera, C. (2005). A kind of auditory 'primitive intelligence' already present at birth. *European Journal of Neuroscience*, *21*, 3201–3204.

Cartwright, T. A., & Brent, M. R. (1997). Syntactic categorization in early language acquisition: formalizing the role of distributional analysis. *Cogntion*, *63*, 121–170.

Chambers, K., Onishi, K., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, *87*(2), B69–77.

Chang, F., Lieven, E., & Tomasello, M. (under review). Word order prediction accuracy.

Charniak, E. (1993). *Statistical language learning.* Cambridge, MA: MIT Press.

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

Chomsky, N. (1959). A review of B. F. Skinner's Verbal Behavior. *Language*, *35*, 26–58.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT

Press.

Chomsky, N. (1970). Remarks on nominalization. In R. Jacobs & P. Rosenbaum (Eds.), *Readings in English Transformational Grammar* (pp. 184–221). Waltham: Ginn.

Chomsky, N. (1980). On cogntive structures and their development: a reply to Piaget. In M. Piatelli-Palmarini (Ed.), Language and Learning: The Debate between Jean Piaget and Noam Chomsky (pp. 35–54). Cambridge, MA: Harvard University Press.

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

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

Chomsky, N. (2004). Language and Mind: Current Thoughts on Ancient Problems. In L. Jenkins (Ed.), *Variation and universals in biolinguistics.* Elsevier.

Christiansen, M., Allen, J., & Seidenberg, M. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, *13*, 221–268.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? an empirical study of the bootstrapping of lexical acquisition. *JAS*, *95*, 1570–1580.

Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (in press). Bootstrapping lexical and syntactic acquisition. *L&S*.

Church, K. W., & Hanks, P. (1989). Word association norms, mutual information, and lexicography. In *Proceedings of the 27th annual meeting of the association for computational linguistics* (pp. 76–83). Vancouver, BC: Association for Computational Linguistics.

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: Italian Ministry of Health, Stella Maris Foundation.

Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language*, *63*, 522–543.

Creel, S., L., N. E., & Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*(5), 1119–1130.

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

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

Davis, S., & Kelly, M. (1997). Knowledge of the English noun-verb stress difference by native and nonnative speakers. *JMemL*, *36*, 445–460.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, *298*(5600), 2013–2015.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Meriaux, S., Roche, A., Sigman, M., et al. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences USA*, *103*(38), 14240–14245.

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

Descartes, R. (1637). *Discours de la méthode.* Paris: Gallimard.

Druks, J., & Froud, K. (2002). The syntax of single words: Evidence from a patient with a selective function word reading deficit. *CogN*, *19*(3), 207–244.

Dryer, M. S. (1992). The Greenbergian Word Order Correlations. *Language*, *68*, 81–138.

Dutoit, T. (1997). *An introduction to text-to-speech synthesis.* Dordrecht: Kluwer Academic Publishers.

Eimas, P., Siqueland, E., Jusczyk, P. W., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development (neural networks and connectionist modeling).*

Cambridge, Massachusets: The MIT Press.

Endress, A. D., & Bonatti, L. L. (2006). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*(e-pub ahead of print).

Endress, A. D., Dehaene-Lambertz, G., & Mehler, J. (in press). Perceptual constraints and the learnability of simple grammars. *Cognition*.

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

Fenk-Oczlon, G., & Fenk, A. (2005). Crosslinguistic correlations between size of syllables, number of cases, and adposition order. In G. Fenk-Oczlon & C. Winkler (Eds.), *Sprache und natürlichkeit, gedenkband für Willi Mayerthaler.* Tübingen: Narr.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of the prosodic modifications in mothers' and fathers' speech to preverbal infants. *JChL*, *16*, 477–501.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences USA*, *99*, 15822–15826.

Fiser, J., Scholl, B. J., & Aslin, R. N. (2007). Perceived object trajectories during occlusion constrain visual statistical learning. *Psychological Bulletin and Review*, *14*, 173–178.

Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: clarifications and implications. *Cognition*, *97*(2), 179–210.

Friederici, A. D. (1985). Levels of processing and vocabulary types: evidence from on-line comprehension in normals and agrammatics. *Cognition*, *19*(2), 133–166.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proceedings of the*

National Academy of Sciences USA, *103*(7), 2458–63.

Fukui, N. (1986). *A theory of category projection and its applications.* Unpublished doctoral dissertation, MIT.

Gallistel, R. (1990). *The organization of learning.* Cambridge, MA: MIT Press.

Gallistel, R. (2000). The replacement of general-purpose learning models with adaptively specialized learning modules. In M. Gazzaniga (Ed.), *The Cognitive Neurosciences* (pp. 1179–1191). Cambridge, MA: MIT Press.

Gambell, T., & Yang, C. (2004). Statistics learning and universal grammar: Modeling word segmentation. In W. G. Sakas (Ed.), *COLING 2004 psycho-computational models of human language acquisition* (pp. 51–54). Geneva, Switzerland: COLING.

Gammon, E. (1969). Quantitative approximations to the word. *Tijd.sch. Inst. Toegepaste Linguistiek*, *5*, 43–61.

Gardner, H., & Zurif, E. (1975). *Bee*, but not *be*: Oral reading of single words in aphasia and alexia. *Neuropsychologia*, *13*, 181–190.

Garrett, M. (1975). The analysis of sentence production. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 133–177). New York, NY: Academic Press.

Gerken, L., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *DP*, *26*, 204–216.

Gerken, L., & McIntosh, B. J. (1993). Interplay of function morphemes and prosody in early language. *DP*, *29*(3), 448–457.

Gertner, Y., Baillargeon, R., Marcus, G. F., Fisher, C. L., & Johnson, S. P. (2007). Rule learning in infants with non-linguistic stimuli. *SRCD Biennial Meeting, 29 March–1 April 2007, Boston, MA, USA.*

Gilbert, C., & Wiesel, T. (1990). The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Research*, *30*, 1689–1701.

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

Gleitman, L. R. (1981). Maturational determinants of language growth. *Cognition*, *10*(1-3), 103–114.

Gleitman, L. R., & Landau, B. (1994). *Acquisition of the lexicon.* Cambridge, MA: MIT Press.

Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *JChL*, *11*, 43–79.

Gold, L., & Lauritzen, M. (2002). Neural deactivation explains decreased cerebellar blood flow in response to focal cerebral ischemia or suppressed neocortical function. *Proc Natl Acad Sci U S A*, *99*, 7699–7704.

Gold, M. E. (1967). Language identification in the limit. *Information and Control*, *10*, 447–474.

Golinkoff, R. M., & Hirsh-Pasek, K. (2007). The view from the radical middle. In H. Caunt-Nulton, S. Kulatilake, & I.-h. Woo (Eds.), *Proceedings of the 31st annual boston university conference on language development, vol 1* (pp. 1–25). Somerville, MA: Cascadilla Press.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135.

Gordon, B., & Caramazza, A. (1982). Lexical decisions for open and closed class words: Failure to replicate differential frequency sensitivity. *B&L*, *15*, 143–160.

Gordon, B., & Caramazza, A. (1985). Lexical access and frequency sensitivity: Frequency saturation and open/closed class equivalence. *Cognition*, *16*, 99–120.

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? *Psychological Science*, *18*(3), 254–260.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *JMemL*, *29*(2), 228–244.

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

Greenberg, J. (1963). *Universals of language.* Cambridge, MA: MIT Press.

Guasti, T. (2002). *Language acquisition.* Cambridge, MA: MIT Press.

Harris, Z. (1951). *Methods in Structural Linguistics.* Chicago: University of Chicago Press.

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

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, *78*, B53–64.

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

Hayes, B. (1995). *Metrical stress theory: Principles and case studies.* Chicago: The University of Chicago Press.

Hayes, J., & Clark, H. (1970). Experiments on the segmentation of an artificial speech analogue. In J. Hayes (Ed.), *Cognition and the development of language* (pp. 221–234). New Yok, NY: Wiley.

Herron, D. T., & Bates, E. A. (1997). Sentential and acoustic factors in the recognition of open- and closed-class words. *JMemL*, *37*, 217–239.

Höhle, B., & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science*, *6*(2), 122–127.

Hubel, D., & Wiesel, T. (1959). Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, *148*, 574–91.

Jenkins, L. (Ed.). (2004). *Variation and universals in biolinguistics.* Elsevier.

Julien, M. (2002). *Syntactic heads and word formation.* Oxford: Oxford University Press.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2000). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure.* Amsterdam: John Benjamins.

Jusczyk, P. W. (1999). Narrowing the distance to language: one step at a time. *Journal of Communication Disorders*, *32*, 207–222.

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

Jusczyk, P. W., & Kemler Nelson, D. G. (1996). Syntactic units, prosody, and psychological reality during infancy. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 389–410). Mahwah, NJ: Lawrence Erlbaum Associates.

Kabak, B., & Vogel, I. (2001). The phonological word and stress assignment in turkish. *Phonology*, *18*, 315–360.

Kay, M. (1973). Morphological analysis. In A. Zampolli & N. Calzolari (Eds.), *COLING 1973 Proceedings of the International Conference on Computational Linguistics 2.* Pisa: COLING.

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

Keurs, M. ter, Brown, C. M., & Hagoort, P. (2002). Lexical processing of vocabulary class in patients with Broca's aphasia: an event-related brain potential study on agrammatic comprehension. *Neuropsychologia*, *9*(40), 1547–1561.

Keurs, M. ter, Brown, C. M., Hagoort, P., & Stegeman, D. (1999). Electrophysiological manifestations of open- and closed-class words in patients with Broca's aphasia with agrammatic comprehension: an event-related brain potential study. *Brain*, *122*, 839–854.

Kiefer, F. (1994). *Strukturális magyar nyelvtan. Fonológia* (Vol. 2). Budapest: Akadémiai Kiadó.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*(1), 15–47.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. *Cortex*, *3*, 163–178.

King, J., & Kutas, M. (1998). Neural plasticity in the dynamics of human visual word recognition. *Neuroscience Letters*, *244*(2), 61–64.

Kovács, I. (2000). Human development of perceptual organization. *Vision Research*, *40*, 1301–1310.

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

Lenneberg, E. (1967). *Biological foundations of language.* New York, NY: Wiley.

Levelt, C., & Vijver, R. van de. (1998). Syllable types in cross-linguistic and developmental grammars. *Third Utrecht Biannual Phonology Workshop, Utrecht, the Netherlands.*

Lewis, D. (1986). *On the plurality of worlds.* Oxford: Basil Blackwell.

Longobardi, G. (2001). The structure of DPs: Some principles, parameters and problems. In M. Baltin & C. Collins (Eds.), *The Handbook of Contemporary Syntactic Theory* (pp. 562–603). Malden, MA and Oxford: Blackwell.

Longobardi, G. (2005). A minimalist program for parametric linguistics? In H. Broekhuis, N. Corver, M. Huybregts, U. Kleinhenz, & J. Koster (Eds.), *Organizing Grammar: Linguistic Studies for Henk van Riemsdijk.* Berlin and New York: Mouton de Gruyter.

MacWhinney, B. (1974). *How Hungarian Children Learn to Speak.* Unpublished doctoral dissertation, UCB.

MacWhinney, B. (1975). Pragmatic patterns in child syntax. *Stanford Papers And Reports on Child Language Development*, *10*, 153–165.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk.* Mahwah, NJ: Lawrence Erlbaum Associates.

Mallinson, G., & Blake, B. (1981). *Language typology: Cross-linguistic studies in syntax.* Amsterdam: North Holland.

Marchetto, E., & Bonatti, L. L. (in preparation). Learning words and rules from an artificial speech stream: Developmental differences at a very young age.

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

Marcus, G. F., Fernandes, K., & Johnson, S. P. (in press). Infant rule learning facilitated by speech. *Psychological Science.*

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

Mazuka, R., Igarashi, Y., & Nishikawa, K. (2006). Input for Learning Japanese: RIKEN Japanese Mother-Infant Conversation Corpus. *The Institute of Electronics, Information and Communication Engineers Technical Report*, *16*, 11–15.

Meek, J. (2002). Basic principles of optical imaging and application to the study of infant development. *Developmental Science*, *5*(3), 371–380.

Mehler, J., & Dupoux, E. (1994). *What infants know: The new cognitive science of early development.* Cambridge, MA: Blackwell.

Mehler, J., Dupoux, E., & Segui, J. (1990). Constraining models of lexical access: The onset of word recognition. In G. Altmann (Ed.), *Cognitive models of speech processing: psycholinguistic and computational perspectives* (pp. 236–262). Cambridge, MA: MIT Press.

Mehler, J., Sebastian Gallés, N., & Nespor, M. (2004). Biological foundations of language: language acquisition, cues for parameter setting and the bilingual infant. In M. Gazzaniga (Ed.), *The New Cognitive Neurosciences.* Cambridge, MA: The MIT Press.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *PsychologR*, *63*, 81–97.

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

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.

Moore, D. R. (2002). Auditory development and the role of experience. *British Medical Bulletin*, *63*(1), 171–181.

Moore, D. R., & Jeffrey, G. (1994). Development of auditory and visual systems in the fetus. In G. Thorburn & R. Harding (Eds.), *Textbook of fetal physiology* (pp. 278–286). Oxford: Oxford University Press.

Morgan, J. L., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Morgan, J. L., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning. *CPsy*, *19*, 498–550.

Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior*, *20*, 67–85.

Morgan, J. L., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax.* Mahwah, NJ: Lawrence Erlbaum Associates.

Mori, K., & Moeser, S. D. (1983). The role of syntactic markers and semantic referents in learning an artificial language. *Journal of Verbal Learning and Verbal Behavior*, *22*, 701–718.

Morita, T., Kochiyama, T., Yamada, H., Konishi, Y., Yonekura, Y., Matsumura, M., et al. (2000). Differences in the metabolic response to photic stimulation of the lateral geniculate nucleus and the primary visual cortex of infants: a fmri study. *Neuroscience Research*, *38*, 63–70.

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, *41*(1), 233–243.

Nespor, M., Guasti, M. T., & Christophe, A. (1996). Selecting word order: the rhythmic activation principle. In U. Kleinhenz (Ed.), *Interfaces in phonology* (p. 126). Berlin: Akademie Verlag.

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

Nespor, M., Shukla, M., Avesani, C., Vijver, R. van de, Schraudolf, H., & Donati, C. (under review). Different phrasal prominence realizations in vo and ov languages? *Cognition*.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology.* Dordrecht: Foris.

Neville, H., Mills, D., & Lawson, D. (1992). Fractionating language: different neural subsystems with different sensitive periods. *Cerebral Cortex*, *2*(3), 244–258.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*, 11–28.

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

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. statistical learning of non-adjacent dependencies in a non-human primate. *CPsy*, *49*, 85–117.

Nobre, A., Allison, T., & McCarthy, G. (1994). Word recognition in the human inferior temporal lobe. *Nature*, *372*(6503), 260–263.

Nonaka, Y., Kudo, N., Okanoya, K., & Mizuno, K. (2006). Statistical learning and word segmentation in neonates: an ERP evidence. *XVth Biennial International Conference on Infant Studies, 19–23 June 2006, Kyoto, Japan.*

Ochs, E., & Schieffelin, B. (1984). Language acquisition and socialization: Three developmental stories. In R. Schweder & R. LeVine (Eds.), *Culture theory: Essays on mind, self and emotion* (pp. 276–320).

Olivier, D. (1968). *Stochastic grammars and language acquisition mechanisms.* Unpublished doctoral dissertation, Harvard University.

Osterhout, L., Bersick, M., & McKinnon, R. (1997). Brain potentials elicited by words: word length and frequency predict the latency of an early negativity. *Biological Psychology*, *46*(2), 143–168.

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

Peña, M., Maki, A., Kovačić, 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.

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress 'deafness'. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7.* Berlin: Mouton de Gruyter.

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

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*(4), 707–784.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition*, *95*(2), 201–36.

Pléh, C., & Juhász, L. (1995). Processing of multimorphemic words in Hungarian. *Acta Linguistica Hungarica*, *43*, 211–230.

Pollack, I., & Pickett, J. (1964). Intelligibility of excerpts from fluent speech: Auditory vs. structural content. *Journal of Verbal Learning and Memory*, *3*, 79–84.

Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *TLR*, *19*, 147–150.

Pulvermüller, F., Lutzenberger, W., & Birbaumer, N. (1995). Electrocortical distinction of vocabulary types. *Electroencephalography and Clinical Neurophysiology*, *94*(5), 357370.

Quine, W. v. O. (1953). *From a logical point of view.* Cambridge, MA: Harvard University Press.

Raichle, M., MacLeod, A., Snyder, A., Powers, W., Gusnard, D., & Shulman, G. (2001). A default mode of brain function. *Proc Natl Acad Sci U S A*, *98*(2), 676–682.

Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition*, *2*, 85–115.

Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *Journal of the Acoustic Society of America*, *105*(1), 512–521.

Reber, A. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855–863.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, *22*(4), 425–469.

Réger, Z. (2004). The Hungarian Réger Corpus.

Rizzi, L. (1986). Null objects in italian and the theory of *pro*. *LingI*, *17*, 501–557.

Rizzi, L. (2005). On the grammatical basis of language development: A case study. In G. Cinque & R. S. Kayne (Eds.), *Handbook of comparative syntax.* Oxford: Oxford University Press.

Roach, P., Arnfield, W., Barry, J., Baltova, M., Boldea, A., Fourcin, W., et al. (1996). Babel: An Eastern European Multi-Language Database. In A. Rubio, N. Gallardo, R. Castro, & A. Tejada (Eds.), *Proceedings of the International Conference on Speech and Language Processing* (pp. 371–374). Granada.

Rohde, D., & Plaut, D. (1999). Language acquisition in the absence of explicit negative evidence: how important is starting small? *Cognition*, *72*(1), 67–109.

Rowland, C. F. (2007). Explaining errors in children's questions. *Cognition*, *104*(1), 106–134.

Ruhlen, M. (1975). A guide to the languages of the world. *Language Universals Project.*

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

Saffran, J. R., Johnson, E., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *JMemL*, *35*, 606–621.

Saffran, J. R., Seibel, R., Pollak, S., & Shkolnik, A. (in press). Dog is a dog is a dog: Infant rule learning is not specific to language.

Segui, J., Mehler, J., Frauenfelder, U., & Morton, J. (1982). The word frequency effect and lexical access. *Neuropsychologia*, *20*(6), 615–627.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.

Selkirk, E. (1996). The prosodic structure of function words. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax.* Mahwah, NJ: Lawrence Erlbaum Associates.

Shafer, V., Shucard, D., Shucard, J., & Gerken, L. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *Journal of Speech, Language and Hearing Research*, *41*(4), 874–86.

Shannon, C. (1948). The mathematical theory of communication. *The Bell System Technical Journal*, *27*, 379–423.

Shi, R., Cutler, A., Werker, J., & Cruickshank, M. (2006). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *JAS*, *119*(6), EL61–7.

Shi, R., Werker, J., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy*, *10*(2), 187–198.

Shi, R., & Werker, J. F. (2001). Six-month-old infants' preference for lexical words. *Psychological Science*, *12*(1), 70–75.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, *72*(2), B11–B21.

Shipley, E., Smith, C., & Gleitman, L. R. (1969). A study in the acuisition of language: Free responses to commands. *Language*, *45*, 322–342.

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

Silverberg, N., Vigliocco, G., Insalaco, D., & Garrett, M. (1998). When reading a sentence is easier than reading a "little" word. *Aphasiology*, *12*, 335–357.

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

Stubbs, M. (1995). Collocations and semantic profiles. *Functions of Language*, *2*(1), 23–55.

Sundberg, U. (1998). Segmental and suprasegmental aspects in infant-directed speech. In *Proceedings of FONETIK, 11th Swedish Phonetics Conference* (pp. 52–55).

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *CPsy*, *50*, 86–132.

Tees, R., & Werker, J. (1984). Perceptual flexibility: maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology*, *38*(4), 579–90.

Teinonen, T. (2007). Measuring newborn skills in statistical learning: an erp approach. *Cognitive Neuroscience and Neuro-imaging of Development Workshop, 19-23 March 2007, Paris, France.*

Thiessen, E., & Saffran, J. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *DP*, *39*(4), 706–716.

Thorne, J. (1968). A computer model for the perception of syntactic structure. *Proceedings of the Royal Society of London, Series B*, *171*(1024), 377–386.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, *74*(3), 209–53.

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

Toro, J. M., Bonatti, L. L., Nespor, M., & Mehler, J. (in press). Finding words and rules in a speech stream: functional differences between vowels and consonants. *Psychological Science.*

Toro, J. M., & Trobalon, J. (2005). Statistical computations over a speech stream in a rodent. *Perception and Psychophysics*, *67*(5), 867–875.

Valian, V., & Coulson, S. (1988). Anchor points in language learning: The role of marker frequency. *JMemL*, *27*, 71–86.

Valian, V., & Levitt, A. (1996). Prosody and adults' learning of syntactic structure. *JMemL*, *35*, 497–516.

Villringer, A., & Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences*, *20*(10), 435–42.

Volterra, V. (1976). A few remarks on the use of the past participle in child language. *Journal of Italian Linguistics*, *2*, 149–157.

Volterra, V. (1984). Waiting for the birth of a sibling: The verbal fantasies of a two-year-old boy. In I. Bretherton (Ed.), *Symbolic play.* New York, NY: Academic Press.

Wenzel, R., Wobst, P., Heekeren, H. H., Kwong, K. K., Brandt, S. A., Kohl, M., et al. (2000). Saccadic suppression induces focal hypooxygenation in the occipital cortex. *J Cereb Blood Flow Metab*, *20*(7), 1103–1110.

Werker, J., Gilbert, J., Humphrey, G., & Tees, R. (1981). Developmental aspects of cross-language speech perception. *CD*, *52*, 349–355.

Werker, J., & Tees, R. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, *37*(2), 278–86.

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

Wexler, K. (1998). Very early parameter setting and the unique checking constraint. *Lingua*, *106*, 23–79.

Wolff, J. (1975). An algorithm for the segmentation of an artificial language analogue. *BJP*, *66*(1), 79–90.

Wolff, J. (1977). The discovery of segments in natural language. *BJP*, *68*, 97–106.

Yamada, H., Sadato, N., Konishi, Y., Kimura, K., Tanaka, M., Yonekura, Y., et al. (1997). A rapid metabolic change in infants detected by fmri. *NeuroReport*, *8*, 3775–3778.

Yang, C. D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Science*, *8*, 451–6.

Zipf, G. (1935). *The psycho-biology of language.* New York: Houghton Mifflin.