[This is a final draft. The published version might contain minor changes.]

Mechanisms of language acquisition: imaging and behavioral evidence

Jacques Mehler
Judit Gervain
Ansgar Endress
Mohinish Shukla

SISSA, Trieste, Italy

Correspondence should be addressed to:

Jacques Mehler
SISSA CNS
via Beirut 2-4, 34014, Trieste, ITALY

mehler@sissa.it

Abstract

Mechanisms of language acquisition have mostly been studied in isolation. Here, we review behavioral and imaging evidence concerning the role and the operation of three such mechanisms: statistical learning, rule extraction and perceptual primitives. Statistical learning is a general learning mechanism, found in animals, adults and infants, that tracks the distributional and statistical information in the input. Rule extraction allows the fast mapping of regularities and the positing of generalizations that go beyond actual experience. Perceptual primitives, the least investigated of the three mechanisms, are specific configurations automatically processed and detected as a result of the way perceptual systems function. We present empirical findings suggesting that neither of the three mechanisms alone is sufficient to explain language development. Rather, the most accurate models of language acquisition will probably emerge from the integration of these and other such mechanisms. Such integrative investigations can greatly benefit from recent advances in brain imaging, such as the use of near infrared optical imaging, the developmental applications of which we briefly discuss here.

1. Introduction

The attempt to explain the uniqueness of language is as old as our own cultural memory. Among the great linguists, Panini investigated the structure of Sanskrit nearly 2500 ago. Grammarians, e.g. Spinoza, pursued the exploration of language structure further, and speculated how different phonological categories are used. Descartes and the Port Royal grammarians made specific proposals about the endowment that allows humans to learn natural language. More than a century later, von Humboldt followed in their footsteps. Instead, more recently, Troubetzkoy and the different structuralist schools were taking a more empiricist stance, deriving much of language structure from the distributional information found in natural languages. It was during the 20th century that these major theoretical traditions have developed into rival theories. On the one hand, psychologists were responsible for the popularizations of some of the most radical versions of empiricism, namely, behaviorism and its more sophisticated contemporary versions, such as connectionism (Elman et al. 96). On the other hand, Chomsky (1965, 1980) proposed the most developed characterization of Universal Grammar and the Principles and Parameters theory, which generative grammarians developed to explain how infants acquire the natural language spoken in our surrounds.

Chomsky's main contribution was to provide the first formulation of the type of linguistic theory that is adequate linguistically, psychologically and biologically. Rather than trying to describe normatively the well-formed utterances of a language, he explicitly stated that the aim of a grammatical theory is to offer the underlying formulae that explain why only the utterances that are grammatical will be generated. Indeed, it is possible to show that native speakers of a language know implicitly the underlying structures that are implemented by the grammar. Last but not least, Chomsky explicitly tied the value of a particular linguistic theory to its ability to account for language acquisition, that is, why it is that any uninjured infant, born into the community, will acquire language with great speed and facility; an ability that generally escapes most adults who are trying to acquire a new language.

Interestingly, theories of language acquisition were explored only from a functionalist perspective. The notion that brain mechanisms as studied in cognitive neuroscience could, at one point, become another source of information for attaining a better understanding language acquisition seemed preposterous to many. Yet, our viewpoint is consistent with the notion that if human languages arose due to a unique endowment characteristic of our species, then a cognitive neuroscience approach to this question is likely to enlighten our research. For instance, Neville and Bavelier (1999:96) state that:

> A general hypothesis that may account for the different patterns of plasticity within both vision and language is that systems employing fundamentally different learning mechanisms display different patterns of developmental plasticity. It may be that systems displaying experience-dependent change throughout life—including the topography of sensory maps, […] lexical acquisition […] and the establishment of form, face and object representations […]—rely upon general, associative mechanisms that permit learning and adaptation throughout life. This type of developmental evidence can contribute to fundamental descriptions of the architecture of different cognitive systems.

This was a position that is reminiscent of the one adopted by Eric Lenneberg (1967) in his *Biological Foundations of Language*. Lenneberg reviews whether the claims that the higher generic learning capacity, as suggested by behaviorists such as Skinner (1957), can account for the facts, and concludes that, contrary to "common-sense" accounts, general intelligence is not correlated with language. More recent studies have strengthened Lenneberg's early writings. In particular, Gleitman and her students (Landau and Gleitman 1985) have observed no language acquisition delays in the blind, contrary to what learning theories would suggest. Likewise, Goldin-Meadow and Mylander (1998) have shown that deaf infants raised in a surround that does not afford linguistic input will spontaneously generate a sign language similar to the already existing sign languages deaf communities use. Today, the naturalistic, as well as the genetic and anatomical information that Lenneberg and others have claimed to be essential to gain understanding about language development, is one that is being actively researched. In this chapter, we also argue that an adequate theory of language acquisition needs to take into consideration some of the basic properties of language, namely, *productivity; partial input; and the ability to acquire multiple natural languages simultaneously.*

Briefly, *productivity* refers to the capacity to understand and generate any well-formed sentence in the language if the lexicon is available. We can transform any thought into a sentence of the language, if we are so inclined, even if we have to invent new lexical terms, as is done continually in science. *Partial input* refers to the capacity of humans to learn the language spoken in their milieu on the basis of a limited amount of fragmentary input. Lastly, *the ability to acquire multiple natural languages simultaneously* refers to the young child growing up in a multilingual environment, who is able to create different files for the various languages spoken in the surrounds without suffering interferences, delay or other of the problems that affect adults in similar situations.

In the past, psycholinguists working on language acquisition did not pay sufficient attention to the resilience of the ability to learn language, despite great deficits. More recently, linguists and psycholinguists formulated theories of language acquisition in which learning had little or no role. But, now the pendulum has again shifted and it would be fair to state that during the last two decades attention has been focused on how statistical machines extract regularities embodied in the linguistic input. Such machines are often taken as providing realistic models of how humans converge on the language spoken in their surrounds, see Hayes and Clark (1970) and Rumelhart et al. (1986), but see also Yang (2004). Unfortunately, we often forget that while arbitrary statistical machines might explain, a posteriori, how the properties of the linguistic signals can shape the native speakers' behavior, they do not address the problem of why it is that non-human primates, who often succeed in statistical learning tasks (Hauser et al. 2000), nevertheless fail to learn human languages, even after prolonged exposure to linguistic stimuli.

The evolutionary accounts of how language arose in humans have been a taboo subject for many decades. However, in the last few years there have been several proposals comparing humans to apes (Hauser et al. 2002, Fitch and Hauser 2004, Fitch et al. 2005, Jackendoff and Pinker 2005, Pinker and Jackendoff 2005). For instance, Hauser, Chomsky and Fitch proposed that to understand the evolution of language it is best to split the study of language into the broad language properties that humans share with other animals and the narrow language properties that may only be present in humans. Concretely, the conjecture that Hauser et al. propose to evaluate is that only humans are capable of performing recursive operations. This view has been challenged by Pinker & Jackendoff, who argue that Hauser and colleagues neglect adaptation as the most likely mechanism capable of explaining the evolution of grammar. While admitting the importance of evolutionary explanations and related cross-species comparisons, our own stance is that the study of the biological foundations of language in contemporary humans, for instance through the investigation of genetic language deficits or genetically endowed language abilities in infants, can provide equally relevant evidence about evolutionary issues. Moreover, the study of prelinguistic infants can greatly clarify what the unlearned precursors are, explaining some of the phonological and morpho-syntactic properties of natural languages. In other words, modern techniques make it possible to explore whether the specific abilities to learn in humans are what shapes the form of natural languages. It may seem paradoxical that most of the work presented below is based on the learning of artificial grammars. However, since many of the experiments attempt to explore both infants and adults, simplification of the materials is desirable.

The first section tries to highlight the brain structures that underlie the dispositions to acquire language that are being detected in the neonate. Imaging methods are many and we focus on near infrared spectroscopy (NIRS), also known as optical topography (OT). Next, we present data suggesting that, rule extraction, statistical learning and perceptual primitives intervene in the acquisition and processing of language, and we argue for their integration into comprehensive models.

2. Language dispositions in very young infants: NIRS studies

Behavioral studies of neonates' perception, attention, and learning abilities have relied on demanding methods to obtain the highly informative data base that we now posses. Indeed, we have a fairly good understanding of how the neonate begins to process faces (Pascalis et al. 2002), colors (Bornstein et al. 1976), and aspects of speech (Jusczyk 1985, Mehler et al. 1988). These discoveries are all the more astonishing considering that large numbers of infants had to be discarded from the experiments because of crying, fussing and several other reasons. Non-nutritive sucking, the most widely used method to test neonates, was notorious. Usually, more than half the tested participants failed to complete the experiments. Three-month-olds and older infants are usually tested using a variety of head-, or eye-turning methods. It is, however, difficult or impossible to test neonates with these methods, see (Aslin et al. 1997).

Behavioral investigations continue to be important for the study of infant development, since they are have already provided a large body of replicable data, and methods continue to improve. However, the search for supplementary methods suitable to study behavior and also inform us about the underlying brain mechanisms responsible for the infants' behaviors is under constant development. Moreover, empirical results should be cross-validated using several methodologies. Thus it is not surprising that investigators are trying to expand the panoply of methods that developmental cognitive neuroscientists can use; some of them are exemplified in other chapters (Friederici et al. present volume).

For well over half a century, developmental science has used physiological measures like EEG and ERP for research purposes. More recently researchers have begun using modern functional imaging techniques with very young infants. Notice however, that fMRI is rather noisy and immobility is required to obtain data, which render this methodology quite difficult to use with young infants.

Nevertheless, a number of studies have been reported to explore the onset of language learning. For instance, some highly informative fMRI studies of language processing have been conducted with three-month-olds (Dehaene-Lambertz et al. 2002, 2006). The first study compared the processing of normal and reversed speech in three-month-olds uncovering a left hemisphere (LH) advantage in temporal areas and in the angular gyrus, much like we observe in adults. Likewise, in the second study with the same age group the authors explored the temporal sequence of activations taking place in different brain areas. The participant infants listened to utterances in their native language while they were being imaged using event related fMRI. The authors found that Heschel's gyrus was the first locus displaying an increased activity. Some time later, both more posterior and anterior areas, including Broca's area also displayed increase activation (Dehaene-Lambertz et al. 2006).

In this section we focus on recent discoveries made possible by NIRS in the domain of language acquisition in neonates and very young infants. NIRS relies on the differential absorption of near infrared light by brain tissue. Near infrared light incident on the skull is scattered, reflected and absorbed to varying extent by various brain tissues. Changes in intensity between the emitted and the recorded light can be related to neural activity, which produces hemodynamic changes, i.e., an increase in oxy-hemoglobin (oxyHb) and a decrease of the deoxy-hemoglobin (deoxyHB), see (Jobsis 1977, Villringer and Chance 1997, Yamashita et al. 1999, Obrig and Villringer 2003). In fact, the extent to which light is absorbed by a medium depends on the wavelength of the near infrared light. The absorption coefficient is a measure of the relative absorbance of light given a particular medium and the wavelength. Choosing the two optimal wavelengths licenses the simultaneous estimation of changes in both oxyHb and deoxyHb. A number of laboratories have already adopted this technology to study the cognitive neuroscience of language development (e.g., Pena et al. 2003, Taga et al. 2003, Bortfeld et al. 2006).

The silence with which NIRS operates is one of the greatest advantages for students of language acquisition in populations of very young infants. Moreover, movements are less critical, since the fiber optics move with the head of the participant. Unfortunately, NIRS only measures emerging photons in a given part of the head, the quantity of which relates to the functionally triggered hemodynamic response, without providing a good enough characterization of the underlying brain anatomy, because in most cases the optical probes are placed on the head using surface landmarks, such as the vertex or the ears.

NIRS-based experiments, like several of the above mentioned fMRI studies, have observed responses to speech stimulation suggesting that the brains of young infants are already organized into areas with functions similar to those observed in older children or adults. For instance, Pena et al. (2003) have shown that infants' brains respond to normal speech differently than to reversed speech, a result that is in many ways comparable to the above mentioned fMRI study and to a behavioral study (Ramus et al. 2000). There are, however, a number of differences, as well. While the NIRS study tested newborns, three-month-olds were tested in the fMRI study. Furthermore, the newborns were mostly sleeping, while the sleeping babies in the fMRI study failed to show activations in some areas that displayed activity when they were awake. Moreover, the NIRS study found that the channels overlaid on the temporal, perisylvian regions of the LH are significantly more activated than the corresponding channels in the RH for normal compared to reversed speech. A more recent unpublished study (summarized in Shukla 2006) attempted to replicate Peña et al. (2003), using a more sophisticated OT machine. This study found basically the same pattern of results although the evidence in favor of a LH superiority in response to speech was restricted to a few channels. These results mesh well with results reported with deaf infants (Holowka and Pettito 2002).

Other studies have expanded the ages of the infants that OT can track. Indeed, Bortfeld et al. (2006) used a sequence of speech plus visual animation interspersed with only visual animation. These blocks were separated using a blank screen presented in total silence. The authors report activations in L-temporal areas during the speech sequence and in occipital regions during exposure to visual animations.

In an investigation with three-month-olds, Homae and colleagues (2006) found that regions of the right hemisphere become activated when infants processed sentential prosody. The authors used short Japanese sentences from a previous behavioral study (Nazzi et al. 1998) under two conditions. In one condition, the original sentences, which were pronounced normally, were used while in the other condition, infants listen to the same sentences, this time with flattened prosody. The authors report that the infants show bilateral activation to the normal sentences. However, when they compared the activation of the normal sentences to the flattened sentences they reported that the channels with the greatest activation are located in the RH temporal-parietal cortex. However, at the individual level, 15 infants show a greater activation in channel 16 in the RH while 10 infants show greater activation in the homologous LH.

In a yet unpublished experiment, Gervain et al. (submitted) showed, using NIRS, that neonates process a string of structured items differently from an otherwise very similar list of items that contain no detectable structure. The structured list consisted of tri-syllabic sequences with a syllable followed by a pair of identical syllables, in short an ABB grammar. The other grammar contained no repetitions, i.e. had an ABC configuration. The anterior areas of the LH show greater overall activation (as measured by changes in the oxyHB concentration) when the neonates are listening to the ABB grammar as compared to listening to the ABC grammar. Moreover, the difference between ABB and ABC grows during the time-course of the experiment. Indeed, the concentration of oxyHB becomes higher for the ABB grammar towards the second part of the experiment, suggesting that infants build a abstract representations only for the structured grammar. As we shall see below, these results can be interpreted from the perspective of purely symbolic computations, as in (Marcus et al. 1999) or from that of configurational perceptual primitives that favor the salience of repetitions in edge positions (Endress et al. 2005).

3. The interaction of statistics and prosodic structures

Since the early seventies psycholinguists have proposed that distributional properties embodied in natural languages are used to extract words and possibly other structural regularities (Hayes and Clark 1970). Indeed, statistical strategies were proposed for the segmentation of words, based on distributional properties over sub-lexical units like phonemes or syllables (e.g., Brent and Cartwright 1996, Batchelder 2002).

Ten years ago, Saffran and her colleagues reported a stunning result, namely, babies segment an artificial grammar composed of tri-syllabic “words” defined by high transition probabilities[1] (TPs) from one syllable to the next. A TP dip between “words” was the only cue available to the eight-month-old infants to parse the continuous string. Details and other work are reported in Aslin et al. (in this volume). Statistical parsing and/or grouping is observed in the auditory, visual and motor domains and in different species.

The original studies by Saffran, Aslin and others simplified theirs stream by disregarding prosodic cues. Johnson and Jusczyk (2001), however, provided evidence for an interaction between various cues. They reported that English 8-month-olds weigh stress and co-articulatory cues more heavily than statistical cues. More recently, Thiessen and Saffran (2003) pitted TPs against stress patterns in English-learning infants, and found that 7-month-olds group bisyllables according to TPs, so a coherent bisyllable is weak-strong, although in English strong syllables are typically word-initial. In contrast, for 9-month-old infants, the stress cues take precedence, and they consider strong-weak, low-TP bisyllables as coherent. Collectively, the various findings suggest that by 9 months of age, infants utilize and integrate multiple cues to word boundaries. However, stress is not the only cue to prosodic structure in spoken language. Thus, sensitivity to larger prosodic constituents can signal the edges of words. Indeed, Gout, Christophe and Morgan (2004) showed that 10- and 12.5-month-olds do not attempt lexical access on syllable sequences that span *phonological phrases* (see also, Soderstrom, Seidl, Kemler Nelson & Jusczyk, 2003).

In addition, young infants have been shown to use intonational phrases in organizing fluent speech (e.g., Mandel, Jusczyk, & Nelson, 1994). For example, Nazzi, Kemler, Jusczyk, and Jusczyk (2000) showed that 6-month-olds could detect previously heard word sequences in fluent speech only if the sequence did not contain an intonational phrase boundary inside it. Different cues, such as statistics and prosody, are present *simultaneously* in fluent speech. Indeed, several researchers have

---

[1] TP(A->B) = P(AB)/P(A), where A and B are units of language, e.g. segments, syllables etc., AB is the co-occurrence of A and B, and P(X) is the probability of the occurrence of unit X.

examined how various cues might interact in segmenting speech into words.

More recently, we have examined possible *models* for how cues interact in speech segmentation. In particular, we asked how the detection of intonational phrases in fluent speech impacts the extraction of statistical regularities (Shukla 2006, Shukla et al. 2006). In these experiments, adults were exposed to carefully controlled artificial speech streams. In this novel paradigm, distributionally coherent (high-TP) trisyllabic nonce words were placed at different locations with respect to artificially generated (intonational) 'phrases'.

----

insert figure 1

Thus, while some words occurred 'phrase'-internally, others straddled such 'phrases'. We found that, in the absence of prosody all the nonce words are recognized, while in the presence of prosody only the 'phrase'-internal words are subsequently recognized.

These experiments allowed us to ask: do prosodic boundaries *inhibit* the computation of TPs across them? We found this not to be the case. Under certain conditions, participants successfully recalled even the contour-straddling words. Thus, we proposed that distributional information is computed independent of the presence of prosodic break points. Only at a later stage do the two cues interact – prosody acts as a *filter*, disallowing sequences that are aligned with prosodic edges.[2]

4. The interaction of distributional information and linguistic categories

The interplay between domain general mechanisms, such as statistical learning, and representations specific to language is an emerging research area. In particular, research is focusing upon the nature of the unit(s) over which statistics are

[2] What drives the perception of such prosodic edges in fluent speech? It is known that the boundaries of prosodic units are associated with acoustic cues like final lengthening and pitch decline (e.g. Beckman & Pierrehumbert, 1986). Indeed, such cues have also shown to be important in detecting 'phrases' in music. For example, Krumhansl and Jusczyk (1990) used a pause-detection paradigm with 4.5- and 6-month-olds and showed that even the younger infants perceived musical phrases as being defined by a pitch decline-reset at phrase boundaries and by a relatively longer final tone. These results suggest that prosodic contours are marked by acoustic patterns that might not be language specific.
We can thus consider a decline in pitch as a perceptual primitive: a pre-existing capacity that is engaged even in a language-specific task – segmenting fluent speech.

computed by human learners. The original statistical learning experiments used artificial streams in which the transitional probabilities were equally informative between units of different kinds, e.g. syllables, consonants, vowels. In natural languages, however, these units play different roles (Nespor et al. 2003). Moreover, cross-linguistic variation in their relative importance and function is also considerable. Therefore, in order to understand how statistical learning might scale up from artificial grammars to the acquisition of a natural language, it is crucial to investigate how linguistic representations, such as consonants or vowels, constrain the extraction of statistical regularities.

Let us first review the different functions of consonants and vowels, as established by linguistic theory, in order to gain insight into how they might interact with statistical learning. The main generalization, supported by numerous empirical observations (Nespor et al. 2003), claims that consonants tend to carry the lexical meanings of words, while vowels express grammatical and morphological functions. Almost universally, languages have more consonants than vowels. Consequently, consonants allow for greater diversity and can encode more information. Thus, they are more adequate than vowels to subserve the storage of a large number of distinctions, characteristic of the lexicon. Vowels, on the other hand, are less numerous, thus less distinct, and even tend to harmonize in certain languages, like Turkish or Hungarian. Importantly, the domain over which vowels harmonize is larger than just the lexical word, and usually encompasses the morphological and some of the syntactic dependents of a word, as well. More direct evidence for the division of labor hypothesis comes from Semitic languages, in which lexical roots are made up of consonants only, which thus define a basic meaning (*-k-t-b-* is the root of words related with 'writing'), whereas the vowels indicate the morphological features of words.

These linguistic observations had been backed up by results from several other domains of research. In psycholinguistics, it had been established that consonants cue the lexicon more than vowels do. In an experiment, Cutler et al. (2000) found that participants prefer to keep the consonants rather than the vowels constant in nonsense words that allow both the change of a consonant and the change of a vowel to yield an existing lexical item (e.g. *kebra* is more often change into *cobra* than into *zebra*). Studies in language acquisition showed that infants lose the discrimination of non-native vowels earlier than that of non-native consonants (Werker and Tees 1984, Kuhl et al. 1992). Language pathologies also provided evidence for the asymmetry between consonants and vowels. Caramazza et al. (2000) reported a double dissociation between them, evidenced by two aphasic patients, one of whom exhibited selective impairment for consonants, while the other showed impairment for vowels.

If true, the division of labor hypothesis makes rather direct predictions about the selective role of consonants and vowels in statistically based segmentation. Since consonants are claimed to carry lexical meaning, it is not unreasonable to expect that they are preferred over vowels for the purposes of statistical segmentation, one of the main uses of which is to assist word learning. Indeed, in the past years, a considerable body of evidence has accrued, suggesting that statistics might be preferentially computed over consonants, but not over vowels.

The initial investigations yielded mixed results. While Newport and Aslin (2004) found that participants segment with equal ease using statistical information over consonants and vowels, Bonatti et al. (2005) obtained segmentation over consonants only. There are, however, a number of differences between the methodologies and materials used by the two groups, possibly explaining their diverging results. Newport and Aslin (2004), for instance, used only two consonantal and vocalic frames as opposed to the three frames of Bonatti et al (2005). Moreover, the former authors allowed immediate repetitions of the same frame in the familiarization, while the latter ones did not. The smaller number of frames and the repetitions in Newport and Aslin’s (2004) experiments might be partly or even fully responsible for successful segmentation with vowels.

This conclusion has been confirmed by further investigations. Toro et al. (in preparation) have found that different mechanisms operate over consonants and vowels in artificial grammar learning situations. While consonants allowed segmentation, but not generalization, the vowels of the same speech stream readily subserved the extraction of regularities. This was true even when the generalization concerning consonants was made very simple (identity) and the information about it was highly redundant. Unpublished work by Shukla et al. have further shown that such simple generalizations (identity) over vowels were easy to learn for participants, and actually prevailed over consonantal TPs.

Taken together, these results argue for the view that the different cues available in language interact with each other. Specifically, the general learning mechanism of TP computations is constrained in language by the nature of the different types of representations present in the input. Some of these representations, e.g. consonants, readily undergo TP computations, because their linguistic function, i.e. encoding lexical distinctions, is compatible with the output of TP computations, i.e. potential word candidates.

The last two sections have addressed the difficult problem of how such a powerful mechanism as statistical computations interacts with other salient properties of natural languages. We saw that while intonational phrases and statistics interact to disallow the statistical nonce-words that straddle boundaries, prosody cannot suppress the automatic statistical computations. We also saw that consonants are a more suitable category of speech upon which to compute statistical dependencies than are vowels. It is premature to say whether this indicates that speakers utilize the knowledge of their native language, which, in most cases has many more Cs than Vs, to select the former over the latter to carry out the parsing routines. It could be the case that an unlearned disposition in humans results in languages that have more Cs than Vs because they are more learnable and the lexicon of such language leads to improved lexical access routine. We are currently conducting NIRS experiments with neonates and four-mont-olds to clarify which of the above options might be correct.

Above we have illustrated the function of a powerful learning mechanism and how it interacts with other properties of languages. We also saw how categories of speech can constrain which of these mechanisms operates best. Now we are going to illustrate other mechanisms and constraints that play an important role in language acquisition. Indeed, it is conceivable that the properties of natural languages honor the functional characteristics of our perceptual organs and most particularly, audition.

5. Perceptual Primitives

Recent research has uncovered two mechanisms to highlight auditory units regardless of whether they are frequent, or statistically salient. One of these mechanisms is the highlighting of edges and the other is the detection of repetitions. As we shall argue below, neither of these mechanisms seems to result from learning.

Edges of domains in speech may modulate how words are segmented, but they may also determine what kinds of generalizations can be extracted from speech streams. One case in point comes from Peña et al.'s (2002) study showing that the inclusion of subliminal silences between words, in otherwise continuous speech streams such as the ones described above, induces participants to extract generalizations. Peña et al. (2002) familiarized participants with a sequence of non-sense words in which the first syllable always predicted the last one, while the middle syllable was variable. The predictive relation between the first and the last syllable could be used in at least two ways. On the one hand, participants could use this relation as a cue to word boundaries, and use this statistical relation to segment the speech stream into its constituent words. Peña et al. (2002) showed that participants do indeed have this capacity. On the other hand, participants may also generalize this relation to new items; in this case, they should accept items as legal if they conform to the dependencies between the first and the last syllables, although they have a different middle syllable. After a familiarization with a continuous speech stream, participants did not accept these generalizations, even when familiarized with a stream of 30 min. However, when words were separated by subliminal silences, a 2 min familiarization was sufficient for inducing the generalizations. Indeed, participants preferred items that had never occurred during the speech stream but that respected the configuration of the edge syllables.[3]

What are the mechanisms underlying this generalization? To address this issue, Endress and Mehler (under review) used pentasyllabic words and asked whether participants would learn generalizations only when the crucial syllables were in the edge positions (that is, the first and the fifth one), or also in middle positions (the second and fourth one). When the critical syllables were in edges, participants readily learned to generalize. In contrast, when the critical syllables were in non-edge positions, participants showed no evidence for the generalizations. Unlike the generalizations, however, statistical processes worked well also in middles. The latter results also suggest that the edge advantage for the generalizations cannot be explained only in terms of the salience of the edges. If it were so, one would expect also statistical processes to break down in middles, which, in fact, they do not do. Hence, edges seem to play a different role for generalization than merely to highlight particular syllables.

Another case in point for the importance of edges in artificial grammar learning has come from phonotactic generalizations. Languages differ in their permissible sound sequences; for example, most consonant clusters would be illegal

[3] More recent research has shown that participants actually extract a regularity entailing syllable classes from such subliminally segmented speech streams, see Endress and Bonatti (in press). The authors showed that participants learn that the first and the last syllable of words have to belong to different syllable classes (the classes being the sets of syllables that can occur in these positions).

in Japanese, but frequent in Polish. Chambers, Onishi and Fisher (2003) showed that young infants can learn constraints on permissible sound patterns from very short exposure. They familiarized participants with CVC (consonant-vowel-consonant) words in which they restricted the consonants that could occur in the first or the last position, respectively. In other words, the first and the last consonants had to come from two distinct sets. After such a familiarization, the infants applied the constraints to new words, thus generalizing them to new instances.

Again, the crucial consonants were placed in the edges of words. To ask whether this feature of the experiments was crucial to the generalizations, Endress and Mehler (under review) asked whether adults could learn similar constraints in longer CVCCVC words. Again, participants had to learn that two consonants had to come from two distinct sets. However, the crucial consonants were in the edges (that is, the first and the last one) for half of the participants, and in middles of the other participants (that is, the second and the third consonant). Participants readily generalized the constraints when the crucial consonants were in the edges, but not when they were in middles. This may be because participants simply do not perceive middle consonants well. However, Endress and Mehler (under review) also showed that participants can discriminate words perfectly well that differ only in their middle consonants; hence, a global impairment for processing middle consonants is unlikely to be the only explanation of the edge advantage for generalizations.

The importance of edges for generalizations in artificial grammar learning can also be demonstrated by considering the experiments by Marcus et al. (1999). In their experiments, young infants were familiarized with syllable sequences conforming to one of the grammars ABA, AAB or ABB (e.g., a sequences like "wo-fe-fe'" would conform to ABB). The infants generalized these grammars to new syllables they have not heard before. While Marcus et al. (1999) argued that these generalizations were evidence for algebraic-like rules even in very young infants, several observations question whether such a claim is justified. First, their structures used repetitions, and we have argued elsewhere that repetition-based structures may be generalized by a simple, specialized operation rather than by a more general rule-extraction mechanism. Second, the repetitions also occurred in sequence edges. To test the role of the edges in this context, Endress, Scholl and Mehler (2005) used 7-syllable sequences (rather than the triplets in Marcus et al.'s (1999) experiments) to ask whether repetition-based structures would be generalized as easily in edges as in middles. They showed that participants generalize repetition-based grammars much more readily when the critical syllables were in edges than when they were in middles; for example they readily generalized the structure ABCDE*FF*, but they failed to generalize the structure ABC*DD*EF. One may be tempted to attribute this result to perceptual difficulties for processing middle syllables. Endress et al.'s (2005) control experiments show that this explanation of the edge advantage is unlikely. They asked participants to *discriminate* sequences that differed either only in middles or in edges; participants still had to process middle (or edge) syllables, but were no longer required to abstract the underlying structure. Participants could discriminate both types of stimuli well above chance. These results suggest that the generalization of such grammars is constrained independently of psychophysical problems for processing middle syllables.

A plausible explanation is thus that only edges have proper positional codes, while other positions can be encoded only relative to such anchor points. One may ask whether the biases reviewed above may be useful also for linguistic phenomena, or only for artificial grammars. While learning syntax obviously entails much more than an edge-detector, even such an operation may be important for some aspects of grammar. The location of words stress in phonology is a first example. Word stress is located relative to either the left or to the right edge; it may be initial or final, or, otherwise, on a syllable counted from the right edge. In contrast, no language has been observed that appeals to word middles, e.g. by locating stress on the middle syllable (e.g., Halle & Vergnaud 1987; Hayes 1995). Morphology also often appeals to edges. Suffixes and prefixes have been observed in many languages, while infixes are rare across languages (e.g., Greenberg 1957).

Another important function of edges may be to interface different levels of representation. For example, morphosyntactic and phonological representations are both hierarchical, but have distinct hierarchies; for example, some morphemes. In such cases, the constituents of the two hierarchies do not coincide; however, at least one of the edges of the constituents must be aligned (Nespor & Vogel 1986; McCarthy & Prince 1993). Edges thus seem to help integrating different hierarchies and levels of representations, and to coordinate them. Surprisingly, mechanisms as simple as an "edge-detector" may thus be important for hierarchical processing, a property that has been considered as crucial for human cognition (e.g., Fodor 1983; Gallistel 1990, 2000; Marr 1982; Marr and Nishihara, 1992). It also highlights that some perceptual biases may have been recruited by the language faculty both for word learning, and for more abstract, structural computations.

6. Discussion and Conclusion

Above we have attempted to show that theorists who focus on one mechanism to the detriment of other mechanisms with which the first interacts may limit our understanding of development. Since Saffran et al (1996), it is recognized that infants rely on distributional cues to segment speech streams. In section 3, we present data that corroborate the importance and automaticity of statistical computations during speech processing. However, we also show that when other sources of information are made available in the input, complementary mechanisms provide a complete processing account. This suggests that studies of language acquisition, while relying on past discoveries must also understand how different processing components mesh with one another, helping to elaborate more naturalistic explorations of language acquisition.

We believe that working with artificial grammars will still prove very useful. However, the more we succeed in scaling up to naturalistic stimuli the more we are going to learn. For instance, consider the ability of neonates to respond differently to the ABB as compared to the ABC grammar, as described earlier. Clearly, making the grammars more complex generates richer models, yielding testable predictions. For instance, comparing grammars containing adjacent repetitions to others with non-adjacent repetitions might instruct us about how working memory develops during the first months of life. That is, making repetitions more and more distant, memory span can be tested. Likewise, introducing prosody in a grammar learning context may also allow us to track in greater detail which cues intervene to constrain the underlying computations.

We also want to stress the importance of perceptual primitives reviewed in Section 5. Kimball (1973) and Bever (1970) claim that perceptual processes are essential to our understanding of how the language user parses novel sentences. Indeed, psycholinguists have experimentally documented the reality of several such claims. Nevertheless, the influence of perception on language acquisition has only recently turned into an active research area. Above, in section 5, we presented research suggesting that repetitions are detected through a primitive identity detector. Gervain et al. (submitted) showed that even newborn infants detect adjacently repeated syllables. Furthermore, Shukla (2006) has demonstrated that the closer a word reoccurs the more it is highlighted. Lindblom and Lacerda (personal communication) have shown that motherese across many languages contains an unsuspected number of word repetitions. Endress has argued that edges of items such as words, phrases or sentences tend to be far more salient than middles. These and other such perceptual primitives should not be ignored. Indeed, those primitives are well documented in the domain of auditory sequential processing. Endress et al. (in press) showed that repetition detectors function with tones, as well as syllables. Whether such perceptual primitives can be attested for visual simultaneous or sequential processing is still an open question.

In brief, we do not think that generalizations, statistics or perceptual primitives should be considered as singletons. Rather we believe that the language acquisition device (LAD) uses all these mechanisms to make language learnable by humans. While Chomsky (1975) formulated the LAD as a framework within which language learning ought to be conceived, the time is ripe to fill in the details giving an outline of how each of the mechanisms fulfills their pre-specified roles. Even the most detailed linguistic theory of how language might be acquired, the Principles and Parameter theory, will ultimately be judged by how well it can integrate all the above mechanisms to explain how an infant goes from signals to abstract grammatical representations.

Bootstrapping theories of language acquisition (e.g. Morgan and Demuth 1996) have isolated some perceptual properties in the speech signal that correlate with abstract grammatical properties. For instance, Nespor (1995) and Nespor et al. (1996) have argued that OV and VO languages place the prosodic prominence at opposite edges of phrases. If so, abstract properties of grammar might be signaled by the prosodic structure of the linguistic data. Since there exist numerous languages that have both OV and VO constructions, it is possible that the frequency of these constructions, together with prosody might select some grammatical properties for a particular language. Notice, however, that the prosodic bootstrapping hypothesis requires that the infant be already endowed with alternative possible grammars ('parameters'). Some properties might arise from the signal plus constraints proper to the perceptual mechanisms of the modality through which language is transmitted.

Thus, in general, prefixing and suffixing are far more frequent as morphological positions compared to infixing. This might arise from the salience of edges in auditory signals. Likewise, grammatical markers tend to appear in edges rather than in middles of constituents.

In conclusion, we have argued in favor of a linguistically informed cognitive neuroscience model of language acquisition. Although we have mostly presented data concerning very basic processes, we have done so considering that the human mind is endowed with the specific disposition to acquire a grammatical system with its appropriate categories. The details of how the human endowment interfaces with the psychological mechanisms that go from universal grammar to particular grammars is still a matter of active investigation.

Last but not least, the progress achieved over the past decade or two in brain imaging has made it possible to explore the endowment for grammar from birth through the first year of life with a facility that was previously unimaginable. Our understanding of the mature brain is constantly increasing making it possible to view the infant’s brain from a perspective of greater ontogenetic continuity than our predecessors had fathomed.

References

Aslin, R. N., Jusczyk, P. W. and Pisoni, D. B. in Handbook of Child Pschology (ed. Siegler, D. K. R.) 147-198 (NY, NY: Wiley, 1997).

Beckman, M. E., Pierrehumbert, J. B. (1986). Intonational Structure in Japanese and English. *Phonology Yearbook* 3: 255--309.

Batchelder, E. O. (2002). Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition* 83(2): 167-206.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In: J.R. Hayes (ed.): *Cognition and the Development of Language*. New York: Wiley, 279-362.

Bonatti, L., Peña, M., Nespor, M., and Mehler, J. (2005). Linguistic Constraints on Statistical Computations: The Role of Consonants and Vowels in Continuous Speech Processing. *Psychological Science*, 16(6): 451 - 459.

Bornstein, M. H., Kessen, W., Weiskopf, S. (1976). Color Vision and Hue Categorization in Young Human Infants. *Journal of Experimental Psychology* 2:115-19.

Bortfeld, H., Wruck, E., Boas, D.A. (2006). Assessing infants' cortical response to speech using near-infrared spectroscopy. *NeuroImage*. 34(1): 407-415.

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

Caramazza, A., Chialant, D., Capasso, R., and Miceli G. (2000). Separable processing of consonants and vowels. Nature 403, 159-160.

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

Chomsky, N. (1975). *Reflections on Language*. New York: Pantheon Books.

Chomsky, N. Aspects of the theory of syntax. (MIT Press, Cambridge, 1965).

Chomsky, N. Knowledge of Language: Its Nature Origine and Use (Praeger New York., 1986).

Chomsky, N. Rules and Representations (Columbia University Press, N.Y., 1980).

Conrad, R. (1960). Serial order intrusions in immediate memory. Br J Psychol, 51, 45-8.

Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., and Van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition, 28* (5), 746-755

Dehaene-Lambertz, G. et al. (2006). Functional segregation of cortical language areas by sentence repetition. *Human Brain Mapping*, 27(5): 360-371.

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

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. *J Exp Psychol Gen*, 134 (3), 406-19.

Endress, A.D. & Bonatti, L.L. (in press). Rapid learning of syllable classesfrom a perceptually continuous speech stream. *Cognition.*

Endress, A.D. & Mehler, J. (under review). Perceptual Constraints in Phonotactic Learning. *J Mem Lang*.

Fitch, W. T. & Hauser, M. D. Computational constraints on syntactic processing in a nonhuman primate. Science 303, 377-80 (2004).

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

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

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

Gallistel, C. (2000). The replacement of general-purpose learning models with adaptively specialized learning modules. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (2nd ed., pp. 1179–91). Cambridge, MA: MIT Press.

Gervain, J., Cogoi, S., Macagno, F., Mehler, J. (submitted). The neonate brain shows a signature for pattern extraction. *Nature*.

Gillette, J., Gleitman, H., Gleitman, L. R., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73(2), 135-76.

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

Gout, A., Christophe, A. & Morgan, J. (2004). Phonological phrase boundaries constrain lexical access: II. Infant data. *Journal of Memory and Language, 51,* 547-567

Greenberg, J.H. (1957) *Essays in Linguistics*. Chicago: University of Chicago Press.

Halle, M. and J.-R. Vergnaud (1987) *An Essay on Stress*. Cambridge, Mass. MIT Press.

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

Hauser, M. D., Chomsky, N. & Fitch, W. T. The faculty of language: what is it, who has it, and how did it evolve? Science 298, 1569-79 (2002).

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

Hayes, J. R. & Clark, H. H. in Cognition and the development of language (ed. Hayes, J. R.) 221-234 (Wiley New York, 1970).

Henson, R. (1998). Short-term memory for serial order: The Start-End Model. *Cognit Psychol*, 36 (2), 73-137.

Henson, R. (1999). Positional information in short-term memory: Relative or absolute? *Mem Cognit*, 27 (5), 915-27.

Hicks, R., Hakes, D., & Young, R. (1966). Generalization of serial position in rote serial learning. *J Exp Psychol*, 71 (6), 916-7.

Hitch, G.J., Burgess, N., Towse, J.N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *Quarterly Journal of Experimental Psychology*. 49A(1), 116-139.

Holowka, S. & Petitto, L. A. Left Hemisphere Cerebral Specialization for Babies While Babbling. Science 297, 1515- (2002).

Homae, F., Watanabe, H., Nakano, T., Asakawa, K., Taga, G. (2006). The right hemisphere of sleeping infants percieves sentential prosody. *Neurosci Research* 54(4): 276-280.

Jackendoff, R. & Pinker, S. The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky) Cognition 97, 211-225 (2005).

Jobsis, F. (1977). Non-invasive infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters *Science* 198: 1264–1267.

Johnson, E. K., Jusczyk, P.W. (2001). Word Segmentation by 8-Month-Olds: When Speech Cues Count More Than Statistics. *Journal of Memory and Language* 44: 548-567.

Jusczyk, P. W. in Neonate Cognition: Beyond the Blooming, Buzzing Confusion (eds. Mehler, J. & Fox, R.) 199-229 (Hillsdale, NJ Erlbaum, 1985).

Jusczyk, P.W., On characterizing the development of speech perception, in Neonate Cognition: Beyond the Blooming, Buzzing Confusion J. Mehler and R. Fox, Editors. 1985, Hillsdale, NJ Erlbaum. p. 199-229.

Kager, Rene (1995) Consequences of Catalexis. in: H. van der Hulst and J. van de Weijer (eds.) *Leiden in Last: HIL Phonology Papers I*. The Hague: Holland Academic Graphics, 269-298.

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

Krumhansl, C. L., and Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science*, 1: 70-73.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255: 606-608.

Landau, B., & Gleitman, L. R. (1985). *Language and experience: Evidence from the blind child.* Cambridge, MA: Harvard University Press.

Lenneberg, E. H. (1967). The Biological Foundations of Language. New York, NY: Wiley.

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

Marr, D. & Nishihara, H. K. (1992). Visual information processing : Artificial intelligence and the sensorium of sight. In S. M. Kosslyn & R. A. Andersen (eds.), *Frontiers in cognitive neurosceince* (pp. 165–186). Cambridge, MA : MIT Press.

Marr, D. (1982). *Vision*. San Francisco, CA: W.H. Freeman.

McCarthy, J. J. & Prince, A. (1993). Generalized alignment. In G. Booij & J. van Marle (Eds.), Yearbook of morphology 1993 (pp. 79–153). Boston, MA: Kluwer.

Mehler, J. et al. (1988). A precursor of language acquisition in young infants. *Cognition* 29: 143-78.

Morgan, J., Demuth, K. (1996). *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Nazzi, T., Kemler Nelson, D., Jusczyk, P., and Jusczyk, A. M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy*, *1*, 123–147.

Nazzi, T., Bertoncini, J., Mehler, J., (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *J. Exp. Psychol. Hum. Percept. Perform*. 24: 756–766.

Nespor, M., Mehler, J., Peña, M. (2003). On the different role of vowels and consonants in language processing and language acquisition. *Lingue e Linguaggio*. 221-247.

Nespor, M. (1995). The role of the phonological phrase for syntactic access. *Proceedings of ABRALIN* 1994. 97- 100.

Nespor, M., Guasti, M. T., Christophe, A. (1996). Selecting word order. In: U. Kleinhenz (ed.): *Interfaces in phonology.* Berlin: Akademie Verlag, 1-26.

Neville, H.J., Bavelier, D. (1999). Specificity and plasticity in neurocognitive development in humans. In M. Gazzaniga (ed.): *The new cognitive neurosciences, 2nd Ed.,* Cambridge, MA: MIT Press, pp. 83-98.

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

Ng, H. L., & Maybery, M. T. (2002). Grouping in short-term verbal memory: Is position coded temporally? Quarterly Journal of Experimental Psychology: Section A, 55 (2), 391?424.

Obrig, H., Villringer, A. (2003). Beyond the visible--imaging the human brain with light. *J Cereb Blood Flow Metab* 23(1): 1-18.

Pascalis, O., de Haan, M., and Nelson, C. A. (2002). Is face processing species-specific during the first year of life? *Science* 296: 1321-3.

Peña, M., Bonatti, L. L., Nespor, M. & Mehler, J. (2002). Signal-driven computations in speech processing. Science, 298(5593), 604-7.

Peña, M., et al. (2003). Sounds and silence: an optical topography study of language recognition at birth. *PNAS* 100(20): 11702-5.

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

Ramus, F., Hauser, M. D., Miller, C., Morris, D. & Mehler, J. Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* 288, 349-51 (2000).

Rumelhart, D. E., McClelland, J. L. & Group, P. R. Parallel distributed processing: Explorations in the microstructure of cognition (MIT Press, Cambridge, MA, 1986).

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Dev Psychol*, 39 (3), 484-94.

Schulz, R. W. (1955). Generalization of serial position in rote serial learning. *J Exp Psychol*, 49 (4), 267-72.

Shukla, M. (2006). *Prosodic constraints on statistical strategies in segmenting fluent speech*. PhD Dissertation, SISSA.

Shukla, M., Nespor, M., Mehler, J. (2006). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*.

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

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

Taga, G., Asakawa, K., Maki, A., Konishi, Y., Koizumi, H. (2003). Brain imaging in awake infants by near-infrared optical topography. *PNAS* 100: 10722-10727.

Thiessen, E. D. & Saffran, J. R. When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. Dev Psychol 39, 706-16 (2003).

Toro, J.M., Bonatti, L., Nespor, M., and Mehler, J. (in preparation). When statistics are computed over consonants and generalizations made over vowles: further asymmetries during language processing.

Villringer, A., Chance, B. (1997). Noninvasive optical spectroscopy and imaging of human brain function, *Trends in Neuroscience* 20, 435-442.

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

Yamashita, Y., Maki, A., Koizumi, H. (1999). Measurement system for noninvasive dynamic optical topography. *Journal of Biomedical Optics, 4(4)*, 414 - 417.

Yang, C. D. Universal Grammar, statistics or both? *Trends Cogn Sci* 8, 451-6 (2004).