[This is a final draft. The published book chapter might include minor changes.]

From perception to grammar

Jacques Mehler*, Ansgar Endress*, Judit Gervain*, Marina Nespor#

*Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste, Italy
#University of Ferrara, Ferrara, Italy

Corresponding author:

Jacques Mehler

SISSA CNS
via Beirut 2-4
34014
Trieste
ITALY

e-mail: mehler@sissa.it
tel.:   +390403787612
fax: +390403787615

1. Biological perspectives on language acquisition

The early writings of Chomsky (1957, 1959), and Lenneberg's *The biological Foundations of Language* (1967) are two examples of how a biological perspective should be incorporated in the explanation of how language is acquired and why other higher vertebrates do not acquire grammatical systems. In this chapter we defend the view that neither of the two favorite views of language acquisition - the "all rule learning" or the "all distributional regularity extraction" – are explanatory when one is conceived without the other. Moreover, as we discuss below, we have worked extensively on another mechanism that we call "perceptual primitive", complementing the former two. By "perceptual primitives", we mean more than just basic perceptual mechanisms that transduce to the brain the stimuli reaching the sensorium. Rather, we try to capture those Gestalt-like organizations of elements or the natural highlighting of certain properties, which then determine many of the properties that can influence or be used by the other two computational mechanisms. In the later sections, we explain how we conceive of these three mechanisms.

Before we do that, let us review just a few salient properties of language that an adequate theory of language acquisition needs to explain:

•*Productivity*: "there are indefinitely many propositions the system can encode" (Fodor & Pylyshyn, 1988). Given knowledge of the lexicon, humans can comprehend any sentence in their language; even those never heard before and can produce an equivalent sentence whenever the thought process renders it necessary. That is, human grammar is indefinitely productive, but, crucially, it relies on a finite set of structural elements (Chomsky, 1957) to achieve this.

•*Partial input*: Humans can learn language on the basis of partial information. The input learners receive comes without explicit indications of structure. Yet, learners extract the regularities that generated the input sequences (which is in itself remarkable given that any finite set of data can be described by indefinitely many different sets of rules) even under very impoverished conditions, such as when deaf children create their own sign language.

•*Acquiring multiple systems*: Humans can simultaneously learn more than one

---

[1] Soon will it be, that to behold these things
Shall not be grievous, but delightful to thee
As much as nature fashioned thee to feel.
(Dante: *Divine Comedy*, Purgatory, Canto XV; English translation by H.W. Longfellow)

language if the input data obliges them to do so. This requires, firstly, that they discover that the input was generated by two (at least partially) different sets of rules and, secondly, that they consistently process and store the rules from the two sets distinctly. Interestingly, the acquisition of multiple languages follows a particular ontogenetic path. Namely, young learners in the phase of first language acquisition learn several languages with almost equal ease and can achieve roughly the same proficiency as their monolingual peers. However, at later ages, learners have more difficulty picking up a new language and, typically, master it less proficiently. Although there is significant individual variation as to how proficient a second language learner may get, the general pattern that late language learning is not in most cases native-like is unequivocal, and arguably derives from how the faculty of language is implemented biologically in the human phylogenetic endowment and its ontogenetic unfolding.

The above traits of language suggest that a theory of language acquisition will be both computational and biological. Undeniably, some species acquire complex song or vocalization patters that have structures remotely reminiscent of some syntactic constraints. But the range of expressions these allow are poor. In fact, no animal communication system has the flexibility to serve purposes other than those programmed to secure the needs of the species, i.e., collecting food, mating, grooming, and so forth. Obviously, humans also use language to secure the same basic needs. Yet, in addition, they frequently use language to express propositions that have nothing to do with either survival or basic needs. They express their states of mind, their beliefs and their desires. They also have many uses for language which allows us to express theoretical ideas, elaborate abstract constructions that make it possible to expand primitive social and cultural settings into ever more detailed social contracts and laws, science and the arts. This suggests that, to obtain a realistic account of language acquisition, one should acknowledge that our brain/mind is different from that of non-human creatures.

Hauser, Chomsky and Fitch (2002, HCF henceforth) proposed that it is convenient to view language as a collection of two components: the first, which they call the faculty of language in the broad sense (FLB), is a collection of abilities we share with other animals; the second is the faculty of language in the narrow sense (FLN) and it is still a conjecture whether it contains a single or multiple components or conceivably even an empty set. Yet the authors' hypothesis (HCF, 2002:1751) is that it is likely to contain only recursion or more precisely,

> "a computational system that generates internal representations and maps
> them into the sensory-motor interface by the phonological system, and into
> the conceptual-intentional interface by the semantic system…the core
> property of FLN is recursion, attributed to narrow syntax".

HCF believe that their comparative approach will lead to new insights and will generate new hypotheses about how the language faculty came to adopt grammatical systems like those now present in all natural languages. In their view the ability to use recursion is an essential ingredient to explain the grammatical systems used by all natural languages. HFC's thesis might turn out to be correct, although we are not certain that the attempt to demonstrate this (Fitch & Hauser, 2004) is convincing (see below).

In this paper we take the stance that there are two informative ways to explore the

biology of language, namely, a ***synchronic*** and a ***diachronic*** perspective. The diachronic approach is the one HFC espouse; it tries to establish how human faculties arose during evolution from precursors supposedly present also in nonhuman animals.

Putatively, precursors evolved to become parts of the language we observe in present day humans. As students of animal cognition, we can inspect the bolts and notches with which humans have been bestowed in possibly phylogenetically distant animals. Unfortunately, there are many aspects of language that we can observe now but whose precursors (except for the shallowest aspects) do not leave traces, as for instance, speech, proto-languages, etc.

Almost four decades ago, Lenneberg (1967:255) already foresaw the growing popularity, as well as the concomitant pitfalls of this approach:

> There were days when learned treatises on the origin of language were based on nothing more than imagination. The absence of ascertainable facts rendered these essays disreputable early during the rise of empirical sciences. For some time the topic became taboo in respectable scientific circles. But recently it seems to have acquired new probity by adumbration of the speculations with empirical data.

This is why a synchronic perspective may contribute to our understanding of the language faculty. Such a perspective focuses on language acquisition and the neural underpinning of language performance. It conceives of our linguistic capacity as a kind of Chomskyan I(nternal)-language. Like Lenneberg in his seminal work, it considers evidence from a broad variety of sources within the same species, ranging from data about neuronal maturation to properties of the perceptual system to particularities of the respiratory system to evidence from speakers/hearers at all ages to abnormalities in patients with brain lesions or who suffer from developmental impairments. In other words, we think that exploring data without having to rely only on conjectures seems a more promising route to understand how languages are acquired, how the performance apparatus works and how grammar is represented in the brain. In this context, novel brain imaging methods may provide particularly informative data for a better understanding of long-debated issues. For example, Peña et al. (2002) used optical topography (OT) to investigate the question whether the lateralization of language is prior to or the consequence of exposure and acquisition. Testing neonates, they have found a left hemisphere advantage in brain activity for speech stimuli as compared to the same speech stimuli played backwards (impossible or non-language) and to silence, suggesting that this hemispheric advantage may not rely on extensive experience. Although it is not clearly established why backward speech functions as non-linguistic material, most probably it is because the human vocal tract is unable to produce backward speech. This is most obvious for stop consonants, whose production is not direction-independent, i.e. requires first a closure, and a burst-like release of the closure, the opposite order is not possible.

For all those reasons we believe that the synchronic route is more promising although there is no denying that these days the synchronic route is not as popular as the diachronic one.

2. Setting the stage: earlier thoughts on language acquisition

In the last four decades, linguists and cognitive scientists studying language acquisition have made tremendous progress, as we shall see below. Today we tend to forget how hard it was to change the dominant paradigm most psychologists espoused in the first half of the twentieth century. Indeed, the classical picture was that of psychologists who overlooked whether their theories were biologically tenable or not; Skinner went as far as to teach and write that he did not believe that studying the brain was of much use. He used to claim that the best way to conceive of the abbreviation CNS was as meaning "conceptual nervous system".

A variety of theories supported the notion that conditioning of various sorts was the essential mechanism underlying language acquisition. In parallel, psychologists also argued that sensitivity to distributional regularities in the environment remains an essential mechanism to acquire language. Information theorists and structural linguists championed this view. Notice, however, that this manner of describing language acquisition avoids mentioning that only the human mind/brain had the disposition to take advantage of such cues to acquire grammar. Moreover, most psychologists working in the early nineteen twenties viewed lexical learning as being the crux of language acquisition, ignoring syntax, semantics, phonology and morphology. This view is still not uncommon today.

In contrast to classical learning psychologists, contemporary students of language acquisition try to focus on syntax without ignoring either the lexicon or phonology. They explore how the complex and abstract structure of syntax arises in the brain of every healthy child who grows up in an environment in which language is utilized. They examine how infants learn to produce and comprehend the sentences of the language they are exposed to, while other organisms that share many of the cognitive abilities humans also posses fail to do so. Synchronic studies also highlight how important it is to study which impairments do and do not result in problems for the pre-lexical infant. Such research has uncovered conditions (such as Specific Language Impairment; see e.g. Gopnik, 1990; Vargha-Khadem et al., 1995) that specifically affect language acquisition, without (or with rather minor) impairments of other cognitive capacities.

The contemporary view of language acquisition tends to merge theoretical and experimental studies of the problem. The formal and computational models of language acquisition influence how empirical scientists couch their research. Moreover, methods to study infants have made great progress making it possible to use neuroscience-inspired methods to expand our understanding of the cortical mechanisms of the changes we observe in early language acquisition. In particular, cognitive scientists have explored systematically both how the child converges to the basic properties of the target language, and the brain changes that accompany acquisition. For example, when children make production mistakes, these may remain constrained by syntactic possibilities attested in some actual natural language (Baker, 2005). Such findings fit well with Chomsky's proposal that, on the one hand, we are born with Universal Principles and, on the other, that we acquire a large range of natural languages. Crucially, these are not arbitrarily variable, but seem to be organized such that different parametric choices account for

variations among languages.

In this chapter, we outline a model of language acquisition that follows the integrative approach introduced above. We thus propose an account that seeks psychological and biological plausibility, while considering the logical problem of language acquisition in its full complexity.

3. Learning language: rule-based and statistics-based approaches

Inspired by Chomsky's rule-based approach to language, early psycholinguistic work strived to understand how humans learn language. Artificial grammars were devised and presented to participants in order to test their ability to extract underlying regularities and generalize them to novel items. The rationale behind these studies was to investigate the learnability of syntax, conceived of as autonomous from other aspects of language. In other words, the question was whether and, if yes, what structures can be learned "in isolation", i.e. in a situation in which participants are deprived of the usual concomitant cues, such as meaning/reference, prosody etc.

Early work in this tradition (Chomsky & Miller, 1958; Reber 1967, 1969 among several others) suggested that participants, familiarized with letter strings generated by an underlying finite-state grammar, were able to discriminate between grammatical and ungrammatical sequences, both novel, despite the fact that they were not consciously aware of the generative rules responsible for the grammatical sequences. In later work, however, it was suggested that success in these tasks relies on learning about bigrams and trigrams, i.e. immediately adjacent sequences of elements, rather than about the more complex underlying pattern that characterizes the string as a whole (e,g. Cleeremans & McClelland, 1991; Dienes, Broadbent, & Berry, 1991; Kinder, 2000; Kinder & Assmann, 2000). However, Reber (1969) also provided experiments that are immune to such criticisms: in these experiments, participants were again familiarized with consonant strings conforming to a finite state grammar --- but then tested on strings from a *new* consonant "vocabulary". As the consonants during familiarization and during test were distinct, successful classification as grammatical or ungrammatical could not be explained by simple statistical computations on the consonants (e.g., Altmann, Dienes, & Goode, 1995; Brooks & Vokey, 1991; Gómez, Gerken, & Schvaneveldt, 2000; Knowlton & Squire, 1996; Meulemans & van der Linden, 1997; Reber, 1969; Tunney & Altmann, 2001).

Another one of the early influential studies was work by Braine (1963, 1966), who approached the learnability issue from the point of view of the structure of natural languages. Specifically, he conjectured that the universal presence of frequent functors (e.g., articles, prepositions/postpositions, pronouns etc.) in the world's languages is a design feature aimed at facilitating learning. These functional elements, easy to track because of their high frequency and their phonological properties, act as anchor points relative to which content words (e.g., nouns, verbs, adjectives etc.) can be positioned. Braine (1963, 1966) tested this hypothesis with 8-10-year-old children in artificial grammar experiments, and found that participants readily learn the position of a non-frequent element relative to a frequent marker (e.g., first, second after the marker), as

opposed to the absolute position of the element. Moreover, Green (1979) and later Morgan et al. (1987) showed that artificial languages in which there are no such markers, or in which the 'content words' are not contingent upon the markers, which thus become bad predictors of structure, are hard or impossible to learn, while a language with reliable markers is fully learnable.

A large body of subsequent work (for example, Morgan and Newport, 1981, Mori & Moeser, 1983; Valian & Seana, 1988; Valian & Levitt, 1996, among others) was concerned with questions similar to Braine's, i.e. how certain features, especially the presence of function words and the existence of larger constituent units, e.g., phrases, contribute to learning. In addition, some of these studies also asked the question how the organization of language in terms of function and content words interacts with other basic properties, such as reference (Mori & Moeser, 1983) or prosody (Morgan et al., 1987; Valian & Levitt, 1996). They found that these additional features facilitate learning, but are not mandatory for learning to take place, and without the function words, they are not sufficient in themselves to induce structure.

Their differences notwithstanding, the quoted studies all have a common feature. They share the interpretation that participants' success in these artificial grammar learning tasks is attributable to their ability to (implicitly) extract rules from the input, an ability that the authors also believed to underlie first language acquisition.

About a decade ago, a new view emerged, reviving pre-generativist ideas about language acquisition. Proponents of this view argue, firstly, that the input to language learners is richer in (statistical) information than previously argued by the generativists, and, secondly, that learners can use their domain-general learning abilities to pick this information up. In an influential paper, Saffran et al. (1996) showed that participants are able to segment a continuous (artificial) speech stream into its constituent word forms solely on the basis of the statistical information contained in the signal. This proposal is based on the intuition, described among others by Shannon (1948) and Harris (1955), that elements (segments, syllables etc.) building up a larger unit are statistically more coherent than elements across unit-boundaries. For instance, in the expression *pretty baby*, the syllable *pre-* predicts the syllable *–tty* with a greater probability than this latter predicts *ba-*, which is part of another word. Following this idea, Saffran and colleagues (1996) constructed a monotonous, synthetic speech stream by concatenating consonant-vowel syllables in such a way that syllables belonging to the same word predicted each other with a higher transitional probability (TP)[2] than those spanning a word boundary. Specifically, they created four trisyllabic nonsense words (e.g., *bidaku*, *golabu* etc.), which were repeated in random order to make up a 2-minute-long speech stream (*bidakupadotigolabubidaku…*), lacking all phonological or prosodic information about word boundaries. The only cues about the words were the TPs, set to be 1.0 between syllables within a word (e.g., *bi-da*), and 0.33 across word boundaries (e.g., *ku-pa*). This stream served as the material to which 8-month-old infants were familiarized. Following familiarization, infants were tested on whether they were able to discriminate between the trisyllabic words of the language and other trisyllabic sequences (called "part-words") that also appeared in the stream, but contained a syllable pair with low transitional

---

[2]   (Forward) transitional probability is a conditional probability statistics defined as: $TP(A \rightarrow B) = F(AB)/F(A)$, where A and B are syllables, and $F(X)$ is the frequency of element X.

probabilities, i.e. that spanned a word boundary (e.g., *kupado*). Indeed, infants attended longer to part-words than to words, indicating that they could discriminate between the familiar words and the statistically illicit part-words. The authors argued that the computation of statistical measures of coherence, such as transitional probability, is a mechanism that potentially plays a very important role in early language acquisition, especially in word segmentation, since it does not require language-specific knowledge on the part of the learner. It has to be noted, however, that it is not equally useful in all languages. Languages that are predominantly monosyllabic, such as Mandarin or (possibly) child-directed English, might pose a problem, since irrespective of TPs, most syllable boundaries are also word boundaries at the same time (Yang, 2002, also about other cross-linguistic issues about TP-based segmentation).

This initial study was followed by a series of others that aimed at exploring different aspects of the statistical learning (SL) mechanism. Thus, it was shown that SL is not specific to linguistic stimuli, operating equally well over tones (Saffran et al., 1999) or visual patterns (Fiser & Aslin, 2002). Neither is it restricted to humans, since cotton top tamarins and rats also succeed in the segmentation experiments (Hauser et al., 2001; Toro et al., 2005).

More importantly, some recent studies tried to clarify what is the actual role and scope of SL in language acquisition. Several questions were raised, including what distance TPs are computed at (only over adjacent element pairs or also over element pairs at a distance), and whether SL also allow rule extraction and generalization. We will address these problems in turn below.

The question of whether TPs can also be computed between non-adjacent elements arose, because in natural language, the relevant structural dependencies that infants need to learn are not only local as in Saffran et al. (1996), but also distant. Dependencies at a distance are universally present in many syntactic structures: e.g. in a sentence as *The children of my brother are coming tonight*, *are* does not refer to the nearest noun, that is *brother*, but to the subject of the sentence, i.e. *the children*. Similarly, in *John promised his kids to buy a new boat*, the subject of *buy* is not the nearest noun *kids*, but *John*.

Similarly, at the level of morphology, many languages have processes of compound formation or inflection, so called parasynthesis or circumfixing, that simultaneously attaches a prefix and a suffix to a stem: neither the prefix plus the stem nor the stem plus the suffix are existing words. For example *devastate* in English is a verb derived from the adjective *vaste*, but neither *devaste* nor *vastate* exist. Or in Hungarian, the superlative of adjectives is formed by adding the circumfix *leg-Adj-bb*, e.g. *jó* 'good', *legjobb* 'best'

Therefore, Peña et al. (2002) devised an adult experiment similar to Saffran et al.'s (1996), except that TPs were made to be predictive between the first and the last syllables of the trisyllabic nonsense words (AXC, e.g., *puliki*, *puraki*, *pufoki*, where *pu* predicts *ki* with a TP of 1.0). Three such distant syllable pairs were defined (*pu-* X *-li*, *ta-* X *-du*, *be-* X *-ga*), and for each distant pair, the same middle items were used (X: *-li-*, *-ra-*, *-fo-*). Consequently, the adjacent TPs, as well as the TPs across word boundaries were uniformly 0.33, and thus were not informative of word boundaries. Participants were familiarized for 10 minutes with a monotonous, synthetic stream, obtained as a

concatenation of the words. When tested with the words versus part-words of the language, participants succeeded in recognizing the words, indicating that they could keep track of and use non-adjacent TPs in order to establish constituent boundaries in the input.[3]

As mentioned before, another crucial question is whether TP computations, and SL in general, allow generalizations or not. Peña et al. (2002) explored this by testing whether participants who were familiarized with the AXC language learn not only the actually attested AXC words, but also the generalization that A predicts C, whatever X comes in between. To answer this question, they have used the same familiarization stream as in the simple non-adjacent TP experiment described above, but they modified the test items in such a way that the part-words were pitted against what they called "rule-words", i.e. trisyllabic words in which the first and the last syllables were provided by one of the word frames (*pu-* X *-be*) of the language, but the intervening X was a syllable that never appeared in that position during familiarization (it was, for instance, the initial syllable of another frame, e.g. *pubeki*, where *be* was the first syllable of the frame *be…ga*). They have found that participants do not generalize, as evidenced by their lack of preference for the rule-words over the part-words. However, once 25 ms subliminal pauses were inserted between the words during familiarization, solving the task of segmentation for the participants, they readily generalized, choosing rule-words over part-words. The authors interpreted these findings as indications that very subtle differences in the properties of the input might induce different processing mechanisms. When TPs are the only cues available, participants compute them to chunk the speech stream into words. However, when other cues are also present, cutting up the stream into constituents, learners engage in different computations and generalize the AXC pattern.

It later turned out that the generalization participants extract seems to be a class-based dependency: They learn that the first syllable of each word has to belong to one syllable class, and the final syllable to another one (Endress & Bonatti, under review). However, also under this interpretation of Peña et al.'s (2002) results, very subtle cues trigger entirely different processes: When the speech stream is segmented even by subliminal silences, participants generalize the structure of the words; in contrast, when no such segmentation cues are given, participants only perform the statistical computations that allow them to segment the stream and do not show any evidence for generalizations.

The question of rule learning was also addressed by Marcus et al. (1999), although from a slightly different perspective, testing directly for generalization in experiments where SL could not take place at all, given that the items used in the test were all novel tokens, thus no TPs could be computed for them during familiarization. Marcus and colleagues familiarized two groups of 7-month-old babies with synthesized

---

[3] Interestingly, in a very similar experiment, Newport and Aslin (2004) have not found better than chance performance on the same word vs. part-word discrimination task. These authors also used trisyllabic nonsense words defined by a TP of 1.0 between the first and the third syllable, but instead of three word frames and three middle syllables as in Peña et al. (2002), they used five word frames and four intermediate syllables. The resulting adjacent TP were thus 0.25 between the first and the second syllables, and 0.20 between the second and the third syllables. The adjacent TPs over word boundaries were also 0.25. It remains to be clarified what is the exact reason for the difference in performance obtained by the two groups of researchers, but the details of the material might play an important role.

languages in which the trisyllabic "sentences" had an ABA (*ga ti ga*) or an ABB (*ga ti ti*) structure, respectively. Then they tested the infants on sentences that were consistent with their familiarization grammar (e.g., ABB for the ABB group) and on sentences that were inconsistent (e.g., ABA for the ABB group). Crucially, however, the items themselves were all made up of novel syllables (e.g., *wo fe fe*), so the only feature that made the test items consistent or inconsistent with the familiarization material was the underlying structure. The authors found that babies looked longer for the inconsistent items, indicating that they discriminated them from the consistent ones. This implies that they have extracted the underlying pattern. As a control, the authors have also run an experiment in which the structure of the two grammars was more similar, i.e., AAB vs. ABB, both containing immediate repetitions. The rationale was to make sure that the babies did not distinguish the two grammars on the basis of cues that were simpler than the structure of the sentences, such as the sheer presence or absence of a repetition. The results showed discrimination under these conditions, too.

An important contribution to our understanding of how generalizations happen come from a series of artificial grammar experiments done by Gomez and Gerken (1999) with 1-year-old infants. In the first experiment, the authors exposed infants to an artificial language generated by a finite state grammar similar to that of Reber (1967), except that they used word strings ("sentence") pronounced by a female speaker instead of letters strings. After a less than two-minute familiarization with grammatical strings deriving from the grammar (e.g., PEL TAM RUD, VOT PEL PEL JIC, VOT JIC RUD TAM etc.), infants were tested on their discrimination between novel grammatical and ungrammatical sentences (e.g., **VOT** PEL PEL **TAM** and **TAM** JIC RUD **VOT**, respectively). These latter were obtained by interchanging the first and the last word of the grammatical sentences. Infants successfully discriminated the two kinds of sentences, as indicated by their significantly longer looking time to the grammatical strings. In a second experiment, using the same familiarization as before, participants were tested on the same grammatical sentences as before, but this time contrasted with ungrammatical sentences that contained licit words in the initial and final positions, but illicit word pairs in the internal slots (e.g., VOT*RUD*PEL JIC, where * marks transitions not allowed by the grammar). Babies, once again, looked longer at the grammatical strings, suggesting discrimination. As a third step, the authors asked whether infants are able to discriminate between two grammars that used the same vocabulary and the same sentence-initial and sentence-final words, but had different "word orders", i.e. different transitions. Two groups of infants were tested, each familiarized with one of the grammars. Then both of them were tested on sentences generated by both of the grammars. For each group, the sentences produced by the grammar they were not exposed to constituted the "ungrammatical" strings. As before, participants showed evidence of learning by looking longer to sentences deriving from their familiarization grammar. Finally, the authors also tested infants' ability to generalize their knowledge about the grammar by familiarizing them to the grammar using one vocabulary (JED, FIM, TUP, DAK, SOG, e.g. JED FIM FIM FIM TUP was a possible sentence), then testing them on sentences coming from the same grammar, but using a new vocabulary (VOT, PEL, JIC, RUD, TAM, e.g. VOT PEL PEL PEL JIC was a corresponding test sentence). This precludes the use of simple transition probabilities between pairs of words to solve the task. Using the same

procedure as in the previous experiment (except for the change of vocabulary between training and test), the authors found that infants can still tell apart their discrimination grammar from the other one, concluding that these results constitute evidence in favor of the learning of abstract linguistic knowledge not reducible to statistics.

In sum, the bulk of the work about the basic aspects of language acquisition has subscribed to one of two interpretations. Participants' performance is either attributed to rule extraction and generalization, or to statistical learning. Although proponents of both views claim that the two processes are not mutually exclusive (Marcus et al., 1999; Newport & Aslin, 2004), there is disagreement as to how labour is shared between the two during acquisition.

4. A new perspective: perceptual primitives in artificial grammar experiments and language

As mentioned in the introduction, the specificity of language is nowadays studied essentially from a diachronic perspective. For example, researchers compare human computational capacities with those of other animals (mostly primates) to draw conclusions about the origins of different aspects of language. However, a synchronic perspective may also yield important insights into language-specificity. In the last section, we reviewed how simplified artificial grammars have been used to explore the kinds of structures that can be learned from simple exposure to exemplars. We will now suggest that two simple, potentially perceptually based, mechanisms can explain, in conjunction with statistical learning, the findings of most of the previous artificial grammar learning experiments. Then we will argue that these mechanisms may also explain more abstract linguistic observations.

As mentioned before, in artificial grammar learning, participants are presented with sequences of linguistic units that conform to a finite state grammar, and then have to judge whether *new* sequences are grammatical or not. They can judge the grammaticality of sequences even when tested on strings from a *different* consonant "vocabulary" than the one during training (e.g., Altmann, Dienes, & Goode, 1995; Brooks & Vokey, 1991; Gómez, Gerken, & Schvaneveldt, 2000; Knowlton & Squire, 1996; Meulemans & van der Linden, 1997; Reber, 1969; Tunney & Altmann, 2001). It turned out that, while the successful classification with the same vocabulary during familiarization and test can be explained by exclusively statistical cues (e.g., Cleeremans & McClelland, 1991; Dienes, Broadbent, & Berry, 1991; Kinder, 2000; Kinder & Assmann, 2000), a different mechanism is responsible when the vocabulary changes between familiarization and test. Indeed, "loops" in the finite state grammars give rise to characteristic patterns of *repetitions*. For example, if strings containing repetitions such as "MTTVT" or "VXVRXRM" are licensed by a grammar, then the pattern of repetitions will also appear when the grammar is instantiated over a *new* consonant set --- since this repetition pattern is a property of the grammar and not of the "vocabulary." Subsequent research has shown that the transfer depends on these repetition patterns; no transfer occurs when the grammars avoid such repetition patterns (see e.g. Gómez et al., 2000; Tunney & Altmann, 2001; see also Brooks & Vokey, 1991).

11

A repetition-based mechanism may also explain other research. Recall for instance that Marcus et al. (1999) used seven-months olds' capacity to generalize repetition-based structures such as ABA, ABB and AAB to argue that these infants can use symbol-manipulation capacities to infer the structure of the stimuli. However, as the grammars entailed repetitions, the infants may simply have detected the repetition-patterns in the stimuli. Indeed, Endress, Dehaene-Lambertz and Mehler (under review) showed that even adults readily learn repetition-based structures but not other, formally equally simple, structures; in particular, using piano tones to carry the structures, they showed that participants readily generalize the structures ABA and ABB (that is, two repetition-based structures), but that they perform poorly for the structures Low Tone – High Tone – Middle Tone and Middle Tone – High Tone – Low Tone.

A mechanism detecting identity relations can also explain results that have been used to draw strong conclusions about specifically human linguistic computations. In particular, HCF (2002) suggested that "recursion" may be a uniquely human capacity that is at the core of the language faculty.[4] To test this hypothesis, Fitch and Hauser (2004) asked whether humans and monkeys could learn a phrase-structure grammar (requiring "recursion") or a finite state grammar. The finite state grammar entailed an alternation of a female and a male voice, while the phrase structure grammar entailed n syllables pronounced by a male speaker followed by another n syllables pronounced by a female speaker (or vice-versa). Human adults learned both types of grammars, while monkeys learned only the finite state grammars. Recent fMRI results by Friederici et al (2005) have rendered the debate about the status of 'recursion' in human language even more interesting. These authors have found that local structural computations, sufficient for learning a finite-state, but not a phase-structure grammar, recruit the left frontal operculum, while computing hierarchical structures necessary for phrase-structure grammars activates Broca's area. The authors interpret this differential activity as further evidence for the separation of local and hierarchical structures, the latter being localized in a brain area especially developed in humans.

While it is certainly possible to cast these stimuli in terms of phrase-structure grammars versus finite-state grammars, a simpler possibility is to assume that participants simply learned the alternations between male and female voices. This seems indeed to be the case: When exposed to strings conforming to the phrase-structure grammar, most participants do not discriminate between strings with equal numbers of male and female syllables and strings with unequal numbers of male and female syllables (which violate the phrase-structure grammar but conform to the pattern of alternations; see Hochmann et al., under review), and the performance decreases dramatically when the saliency of the alternation is reduced (using a contrast between Consonant-Vowel-Consonant (CVC) and Consonant-Vowel (CV) syllables instead of the male vs. female contrast).

The importance of repetitions has also been acknowledged implicitly by researchers studying statistical learning. Indeed, when preparing speech streams for such experiments, care is always taken to avoid immediate repetitions of words (e.g., Saffran et al., 1996); such repetitions seem to make the words pop out. In line with this

---

[4] "Recursion" is used by HCF in a way that is more general than how this notion is most often understood in linguistics. These authors employ it to refer to embedding in general, and not only to embedding of a constituent in a constituent of the same category.

possibility, speech streams are segmented only if words are "repeated" closely enough, but not when two occurrences of the same word are separated by too many intervening items (M. Shukla, in preparation).

The arguments reviewed above suggest that a mechanism detecting identity relations can explain a wide range of data. Other experiments can be explained by another, equally simple, mechanism. One can illustrate this mechanism again with Marcus et al.'s experiments. In their experiments, the repetitions always occurred in the *edges* of sequences; that is, the repetitions occurred either sequence-initially or sequence-finally. Now, it is known since Ebbinghaus (1885) that edges are salient, and remembered better. (We will discuss below how the edges may favor generalizations.)

The importance of edges in language acquisition has first been stressed in corpus-based studies. For example, grammatical constructions such as auxiliaries or root infinitives are more frequent in child-language if the corresponding constituents appear in sentence-edges (e.g., Gleitman, Newport, & Gleitman, 1984; Wijnen, Kempen, & Gillis, 2001). It is thus possible that the structures in Marcus et al.'s (1999) experiments were particularly easy to extract because the repetitions were in the edges, unlike what Marcus et al.'s (1999) more general interpretation would suggest. Endress, Scholl and Mehler (2005) tested this hypothesis by asking whether participants would generalize repetition-based grammars, where the repetitions were either in sequence-edges (e.g., ABCDE**FF**) or in sequence-middles (e.g., ABC**DD**EF). Indeed, participants readily generalized the edge-repetitions but were close to (or at) chance for the middle-repetitions. Still, the advantage for edge-repetitions was not simply due to participants remembering syllables better in edges than in middles; indeed, when asked to *discriminate* the same syllable sequences (that is, they still had to process the sequences but they had no longer to generalize their structures), participants performed at ceiling for both edge- and middle-repetitions, suggesting that the generalization ability is specifically constrained by where in the sequence the relevant structure occurs.

Edges proved to be important also for other experiments. For example, Chambers, Onishi and Fisher (2003; see also Onishi, Chambers & Fisher, 2002, for experiments with adults) showed that infants can learn phonotactic-like constraints from brief exposure; in particular, the infants learned that, in Consonant-Vowel-Consonant (CVC) words, the initial and the final consonants had to come from distinct sets. As the crucial consonants were the words-edges, one can ask whether this feature was crucial to the generalizations. Later studies observed such phonotactic-like generalizations when the crucial syllables were at word edges, but not when they were in word-middles (Endress & Mehler, under review). Again, a control experiment showed that the edge advantage was not due to a brute impairment for processing consonants in word middles, but that these edge-based constraints seem to affect more the generalization ability than the ability to process consonants.

Yet another example comes from Peña et al.'s (2002) experiments. Recall that these authors showed that when participants are familiarized with a quasi-continuous speech stream, the inclusion of subliminal 25 ms silences between words triggers generalizations that are not available otherwise. Under these conditions, participants seem to extract a class-based dependency: They learn that the first syllable of each word has to belong to one syllable class, and the final syllable to another one (Endress & Bonatti,

13

under review). Again, the crucial syllables occurred in the edges, so one may ask whether this feature of the experiment was crucial to the generalizations. As in the other experiments, the class-based generalizations were observed when the crucial syllables occurred in word edges, but not when they were in the middle syllables (Endress & Mehler, unpublished). It thus seems that the class-based generalizations were triggered by the edges in which the crucial syllables were placed. This explanation accounts also for previous experiments studying how word classes can be learned, as the crucial elements occurred at edges also in these experiments (Braine, 1963, 1966; Smith, 1966, 1967, 1969).

These results suggest that two mechanisms may be important for structures used in Artificial Grammar Learning experiments: an operation sensitive to repetitions, and a second operation specifically attending to word edges. We will discuss now these operations in turn.

First turning to repetitions, Gervain et al. (in preparation) have recently shown, using optical imaging, that a sensitivity to repetitions may arise ontogenetically very early. The authors have found that the neonate brain readily distinguishes between sequences containing repetitions at the edges of items (similar to Marcus et al.'s ABA and ABB sequences) and matched controls that do not contain such repetitions (ABC sequences), as evidenced by larger and longer-lasting activation for the former type of stimuli. Thus, we might hypothesize that edges and repetitions act as perceptual primitives that infants can detect from the very beginning of language acquisition, and might use them as basic building blocks to learn some aspects of more abstract structures.

These results also suggest that identity relations may not be a peculiarity of Artificial Grammar Learning experiments, but rather that they may be more widespread. Indeed, a basic operation sensitive to identity relations may also explain a range of linguistic observations. Reduplication is a widespread phenomenon in morphology that entails the repetition of (part of) a word root (McCarthy & Prince, 1999). It can either create new words, through derivation or composition, or new forms of a word through inflection. Though total reduplication is sometimes attested, partial reduplication is more frequent. For example, in Marshallese, a Malayo-Polynesian language spoken in the Marshall Islands, reduplication of the syllable at the right edge of a word is used to derive verbs from nouns. For example, *takinkin* ('to wear socks') is derived from *takin* ('sock'; Moravcsik, 1978). In Classical Greek, left edge reduplication is used in verbal inflection: λείπω [leipo]'I leave', λέλοιπα [leloipa] 'I left'. Although medial reduplication is attested, it is rare compared to either initial or final reduplication.

Other examples of reduplications, found in languages are rhyming reduplication as in *abracadabra, boogie-woogie, hocus-pocus*, total reduplication as in *bonbon, bye-bye, couscous* or reduplication with a vowel change as in *flip-flop, hippety-hoppety, kitcat, zig zag, ping-pong*.

Edge-based regularities may be even more widespread in linguistics. The data reviewed above is already suggestive of the generality of such phenomena: edge-based phenomena could be observed at sequence edges, sentence edges, word edges, and probably still under other conditions; the crucial items could be phonemes, syllables or words. It thus seems that edge-based regularities may be exploited by natural languages at different levels of description.

14

Before discussing the linguistic phenomena that may appeal to edges, however, it is worthwhile to discuss what the role of edges may actually be. In the first demonstrations of an edge advantage, Ebbinghaus (1885) observed that items in edge positions are remembered better than items in non-edge positions. Later research found that this is not the only role of the edges. Indeed, participants have not only to learn *that* particular items occur in a sequence, but also *where* the items occur. For example, one type of error in sequence recall are sequentially correct *intrusions*, where an intruder is recalled in its correct sequential position but in a sequence it has never appeared in (e.g., Conrad, 1960; Henson, 1998, 1999; Ng & Maybery, 2002). Such results can only be explained if participants learn *positional codes* for each item that are independent of the sequence the item appeared in; such positional codes seem to undergo their own serial position effect, such that it is easier to remember that a particular item occurred in the first or the last position than to remember that it occurred in a middle position (e.g., Conrad, 1960; Henson, 1998, 1999; Hicks, Hakes, & Young, 1966; Ng & Maybery, 2002; and many others). As the edges may be the most reliable positional code, it seems plausible that they may be used by many different processes to define regularities. Indeed, most contemporary models of positional codes in sequences assume, in some form or the other, that *only* the edges have proper positional codes, and that the other positions are encoded with respect to the edges (e.g., Henson, 1998; Hitch, Burgess, Towse, & Culpin, 1996; Ng & Maybery, 2002). Such results also suggest that edges may not be "hard" limits to generalizations, but it is probably possible to draw generalizations "close to" edges and to observe a graceful degradation afterwards[5].

Linguistic regularities extensively appeal to edges. In phonology, for example, word stress rules make reference to either the left or to the right edge. Stress may be initial (e.g., in Hungarian) or final (e.g., in Turkish) or on a different syllable defined starting from the right edge. For example, in Latin (as well as many other languages), stress is defined on the basis of a word's right edge: it is penultimate (i.e. second from the right) if the penultimate syllable is heavy, antepenultimate (i.e., third from the right) otherwise. No language makes reference to the middle of words, e.g. by stating that stress falls in the middle syllable (Halle & Vergnaud, 1987; Hayes, 1995; Kager, 1995). Interestingly, if word stress does not fall at the same position within the word, it is computed from the right, but not from the left edge (Hayes, 1995; Kager, 1995).

Phenomena of phrasal phonology often apply to one edge of a constituent or across two constituents to signal their syntactic cohesion, by eliminating their edges. An example of the first type is the final devoicing of voiced stop consonants in Dutch (be[t] vs. be[d:]en 'bed' vs. 'beds'). An example of the second type is *liaison* in French, the resyllabification of the final consonant of a word with a vowel-initial following word; this has the effect of eliminating the edges that separate the words. It occurs for example between articles and nouns (as in *les enfants*, le[z]enfant 'the children') or between auxiliaries and verbs (as in *je suis allé*, jesui[z]allé 'I have gone'), but it does not apply between a subject and verb (as in *les enfants ont mangé*, les enfants[Ø]ont mangé 'the children have eaten') to signal that the two constituents have a low level of cohesion (Nespor & Vogel, 1986).

---

[5] Also identity may not be an all-or-none relation, but may be the extreme case of similarity of two items (see e.g., Frisch et al., 2001).

The morphological process of affixation also clearly privileges edges: languages are rich in suffixes and prefixes, while infixes are rare (Greenberg, 1957). In addition suffixes are more frequent than prefixes (Sapir, 1921; Dryer, 2005; Cutler et al., 1985; Hawkins & Cutler, 1988); in a cross-linguistic database of grammatical markers 74.4% are suffixes (Bybee et al., 1990). While there are languages, such as Turkish, Basque, Burmese or Hindi, that have only suffixes, language with only prefixes, like Thai, are quite rare (Greenberg 1963). This asymmetry between prefixes and suffixes would seem to suggest, as the phenomenon of stress assignment seen above, that the right edge is perceptually more salient than the left edge.

Edges are not only privileged positions for various types of linguistic processes; they are also crucial for the mapping of different levels of representation. Morphosyntactic and phonological representations are both hierarchical in nature, but the two hierarchies are distinct: while *dis* is a morpheme in *disillusion*, it is not a syllable. Constituents of the two hierarchies often coincide, but when they do not, they are never totally mismatched: at least one of the edges – either left or right – must be aligned. For example, the left edge of a syntactic phrase is aligned with the left edge of a phonological phrase in right recursive languages, that is, languages with subordinate clauses after main clauses and complements after heads (as English or Spanish); in contrast, the right edge is not necessarily so aligned. The reverse is true in left recursive languages, that is, in languages with subordinate clauses before main clauses and complements before heads (as Turkish or Japanese): In such languages, the right edges of the two constituents are necessarily aligned, but not the left edges (Nespor & Vogel, 1986).

For example, the sentence [*John*] [*bought*] [*some nice land*] contains three phonological phrases, as indicated by the brackets. In all three cases the left edge of the phonological phrase is aligned with the left edge of a syntactic phrase: the subject noun phrase, the verb phrase and the object noun phrase. In the first and third phonological phrase also the right edge is so aligned, but not in the second: no syntactic phrase ends after the verb. The opposite is true in a Turkish phrase as [*Mehmet*] [*cam*] [*kIrdI*] (Mehmet – window – broke) 'Mehmet broke a window'. In the two noun phrases, both edges are aligned with the edges of a phonological phrase. Not so for the verb phrase, where only the right edge is aligned. The left edge is not the beginning of the verb phrase, that includes also the object *güzel ev*.

It thus appears that although there is not necessarily a one-to-one correspondence between the constituents of the two hierarchies, at least one of the edges of the two must coincide. This capacity of edges to mediate between different hierarchies and levels of representation is a surprisingly powerful notion for an operation as simple as edges. Indeed, hierarchical processing has long been thought to be a fundamental property of human (and presumably other animals') cognition (e.g., Fodor, 1983; Gallistel, 1990; Gallistel, 2000; Marr, 1982; Marr & Nishihara, 1992); this gives rise to the need for mechanisms by which different levels of representation can be put into correspondence. If different hierarchies independently define their edges, the edges may be in some cases the common currency by which these hierarchies can be coordinated (McCarthy & Prince, 1993; Nespor & Vogel, 1986).

In sum, many Artificial Grammar Learning experiments can be explained by two simple mechanisms: a mechanism sensitive to identity relations, and another one

attending specifically to edges. Both mechanisms seem to be shared by other animals: Non-human primates both are sensitive to positional codes (e.g., Orlov et al., 2000) and generalize identity relations (e.g., Hauser, Weiss, & Marcus, 2002; Wallis, Anderson, & Miller, 2001), a capacity that is shared even with honeybees (Giurfa et al., 2001). Nevertheless, the language faculty seems to use such "perceptual primitives" extensively for its structural computations, which may shed some light on at least parts of its origins using purely synchronic investigations.

5. Conclusion

In this paper, we have presented two views on the specificity of language. While the currently more popular view is based on a diachronic perspective and compares the capacities of different animals, we have suggested that also synchronic observations may yield crucial insights. We have illustrated this approach first by considering a wide variety of experiments in Artificial Grammar Learning. We showed that much of this work can be explained by two simple "perceptual primitives" specifically tuned to certain salient patterns and configurations in the input: An operation sensitive to identity relations, and another operation specifically sensitive to edges. We then reviewed linguistic observations suggesting that the language faculty makes extensive use of these very same perceptual primitives. These primitives suggest a new way in which people may learn from their environment: In addition to ubiquitous statistical mechanisms, such Gestalt-like primitives may allow them to extract particular structures from the input. Both statistical computations and perceptual primitives may then interact and feed into more abstract computations; in this way, they may also contribute to learning parts of morpho-syntax. A provocative possibility is that also more abstract levels of representation have their own versions of such primitives; for example, some edges of linguistic constituents (such as morpheme boundaries) are not also signalled by "acoustic" edges. From purely synchronic investigations, we may thus have identified two psychological mechanisms, namely an operation sensitive to identity relations and one sensitive to edges, that may be used by the language faculty (but were presumably present before language arose). It may thus be possible to understand some linguistic computational principles by considering principles of perceptual organization.

BIBLIOGRAPHY

Altmann, G. T., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *J Exp Psychol Learn Mem Cogn*, 21(4), 899–912.

Baker, M. (2005). The innate Endowment for Language: Overspecified or Underspecified? In P. Carruthers, S. Laurence and S. Stitch (eds.) *The Innate Mind: Structure and Contents.* Oxford University Press, Oxford.

Bavelier, D. & Potter, M. (1992). Visual and phonological codes in repetition blindness. *J Exp Psychol Hum Percept Perform*, 18(1), 134-47.

Bavelier, D. (1994). Repetition blindness between visually different items: the case of pictures and words. *Cognition*, 51(3), 199-236.

Braine, M. (1963). On learning the grammatical order of words. *Psychol Rev,* 70, 323-48.

Braine, M. (1966). Learning the positions of words relative to a marker element. *J Exp Psychol*, 72(4), 532-40.

Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *J Exp Psychol Gen*, 120(3), 316–323.

Bybee, J., W. Pagliuca, and R. D. Perkins (1990). On the asymmetries in the affixation of grammatical material. *Studies in Typology and Diachrony*, ed. by William Croft, 1-42. Amsterdam: Benjamins.

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

Chomsky, N. & Miller, G. (1958). Finite State Languages. *Information and Control* 1: 91-112.

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

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

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. J Exp Psychol Gen, 120(3), 235-53.

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

Cutler, A., J. A. Hawkins, and G. Gilligan (1985). The suffixing preference: a processing explanation. *Linguistics*, 23.723-58.

Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *J Exp Psychol Learn Mem Cogn*, 17(5), 875-87.

Dryer, M. S. (2005) Position of case affixes. *World Atlas of Language Structures*, ed. by M. Haspelmath, M. S. Dryer, D. Gil and B.Comrie, 210-213. Oxford: Oxford University Press.

Ebbinghaus, H. (1885/1913). Memory : A contribution to experimental psychology. New York : Teachers College, Columbia University. (http ://psychclassics.yorku.ca/Ebbinghaus/)

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. (under review). Rapid learning of syllable classesfrom a perceptually continuous speech stream. *Cognition.*

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

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

Fiser, J., & Aslin, R.N. (2002). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning,*

Memory, and Cognition, 28(3), 458-467

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

Fodor, J. A. & Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition*, 28(1-2), 3-71.

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

Friederici, A., Bahlmann, J., Heim, S., Schubotz, R., Anwander, A. (2005). The brain differentiates human and non-human grammars. PNAS, 103(7), 2458-2463.

Frisch, S. A., Pierrehumbert, J. B. & Broe, M. B. (2004). Similarity avoidance and the OCP. *Nat Lang Ling Theory*, 22(1), 179–228.

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., Mehler, J. (in preparation). Ready for structure? The neonate brain's ability to detect linguistic structure.

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of 'sameness' and 'difference' in an insect. *Nature*, 410(6831), 930-3.

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

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

Gómez, R. L., Gerken, L., & Schvaneveldt, R. (2000). The basis of transfer in artificial grammar learning. *Mem Cognit*, 28(2), 253-63.

Gopnik, M. (1990). Feature-blind grammar and dysphasia. *Nature* 344, 715.

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

Greenberg, J.H. (1963) Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In: J.H. Greenberg (ed.). *Universals of Language*. London. MIT Press, pp. 92-96.

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

Harris, Z. (1955). From phoneme to morpheme. *Language* 31.190-222.

Hauser, M.D., Newport, E.L. & Aslin, R.N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition* 78: B53-B64

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), 1569-79.

Hauser, M. D.,Weiss, D. & Marcus, G. (2002). Rule learning by cotton-top tamarins. *Cognition*, 86(1), B15-22.

Hawkins, J. A. and A. Cutler (1988) Psycholinguistic factors in morphological asymmetry. *Explaining Language Universals,* ed. by John A. Hawkins, 280-317. Oxford: Blackwell.

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

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.

Hochmann, J-R, Azadpour, M. & Mehler, J. (submitted). Specific language abilities lost in a methodological labyrinth. *Science*.

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.

Kanwisher, N. & Potter, M. (1989). Repetition blindness: the effects of stimulus modality and spatial displacement. *Mem Cognit*, 17(2), 117- 24.

Kanwisher, N. (1987). Repetition blindness: type recognition without token individuation. *Cognition*, 27(2), 117-43.

Kanwisher, N., Driver, J. & Machado, L. (1995). Spatial repetition blindness is modulated by selective attention to color or shape. *Cognit Psychol*, 29(3), 303-37.

Kinder, A. (2000). The knowledge acquired during artificial grammar learning: Testing the predictions of two connectionist models. *Psychol Res*, 63(2), 95-105.

Kinder, A., & Assmann, A. (2000). Learning artificial grammars: No evidence for the acquisition of rules. *Mem Cognit*, 28(8), 1321-32.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *J Exp Psychol Learn Mem Cogn*, 22(1), 169-81.

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

Lenneberg, E. (1967). *The Biological Foundations of Language.* 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. (Reprinted from Technology Review 81 :2–23 (1978))

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

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.

McCarthy, J. J. & Prince, A.S. (1999). Faithfulness and identity in prosodic morphology. In R. Kager, H. van der Hulst, and W. Zonneveld (Eds.), *The prosody morphology interface*. Cambridge: Cambridge University Press. 218-309.

Meulemans, T., & van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. J Exp Psychol Learn Mem Cogn, 23(4), 1007–1028.

Miller, M. & MacKay, D. (1994). Repetition deafness: Repeated words in computer-

20

compressed speech are difficult to encode and recall. Psychol Sci, 5, 47–51.

Miller, M. & MacKay, D. (1996). Relations between language and memory: The case of repetition deafness. Psychol Sci, 7, 347–51.

Moravcsik, E. (1978). Reduplicative constructions. In J. H. Greenberg (Ed.), *Universals of human language: Word structure*. Stanford, CA: Stanford University Press. 3. 297-334.

Morgan, J. L., & Demuth, K. (Eds.). (1996). Signal to syntax: Bootstrapping from speech to grammar in early acquisition. Mahwah, NJ: Lawrence Erlbaum.

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., Meier, R. P., & Newport, E. L. (1987). Structural packaging in the input to language learning. Cognitive Psychology, 19, 498–550.

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.

Näätänen R, Gaillard AWK and Mantysalo S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol* 42:313-329

Näätänen R, Tervaniemi M, Sussman E, Paavilainen P, Winkler I. (2001). "Primitive intelligence" in the auditory cortex. *Trends Neurosci*. 24(5):283-8.

Nespor, M. and I. Vogel (1986) *Prosodic Phonology*. Foris: Dordrecht.

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.

Onishi, K. H., Chambers, K. E. & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. Cognition, 83(1), B13-23.

Orlov, T., Yakovlev, V., Hochstein, S. & Zohary, E. (2000). Macaque monkeys categorize images by their ordinal number. Nature, 404(6773), 77-80.

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

Peña, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., Mehler, J. (2002). Sounds and silence: An optical topography study of language recognition at birth. PNAS, 100(20), 11702-11705.

Pinker, S. (1984). *Language Learnability and Language Development*. Harvard University Press, Cambridge, MA.

Pinker, S. & Bloom, P. (1990) Natural language and natural selection. *Behavioral and Brain Sciences, 13*, 707-784.

Pinker, S. & Jackendoff, R. (2005). The faculty of language: What's special about it? Cognition, 95, 201-36.

Premack, D. and Premack, A. (2003). *Original Intelligence*. McGraw-Hill, New York.

Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology, 81,* 115-119.

Reber, A. S. (1967). Implicit learning of artificial grammars. J Verb Learn Verb Behav,

6(6), 855–863.

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

Saffran, J., Johnson, E., Aslin, R., & Newport, E. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1), 27-52.

Sapir, E. (1921) *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace & Co.

Shannon, C.E. (1948). *A Mathematical Theory of Communication*, Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656.

Shukla, M. in preparation. *An interaction between prosody and statistics in segmenting fluent speech*. PhD dissertation, SISSA, Trieste.

Smith, K. (1966). Grammatical intrusions in the recall of structured letter pairs: mediated transfer or position learning? J Exp Psychol, 72(4), 580-8.

Smith, K. (1967). Rule-governed intrusions in the free recall of structured letter pairs. J Exp Psychol, 73(1), 162-4.

Smith, K. (1969). Learning co-occurrence restrictions: rule learning or rote learning. J Verb Learn Verb Behav, 8, 319 ½ U-21.

Soto-Faraco, S. (2000). An auditory repetition deficit under low memory load. J Exp Psychol Hum Percept Perform, 26(1), 264-78.

Toro, J.M. & Trobalón, J.B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, 67, 867-875.

Tunney, R., & Altmann, G. T. (2001). Two modes of transfer in artificial grammar learning. J Exp Psychol Learn Mem Cogn, 27(3), 614-39.

Valian, V., & Levitt, A. (1996). Prosody and adults' learning of syntactic structure. Journal of Memory and Language, 35, 497–516.

Valian, V., & Seana, C. (1988). Anchor points in language learning: The role of marker frequency. Journal of Memory and Language, 27, 71–86.

Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P. & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proc. Natal. Acad. Sci. USA* 92, 930−933.

Wallis, J., Anderson, K., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. Nature, 411(6840), 953-6.

Wijnen, F., Kempen, M. & Gillis, S. (2001). Root infinitives in dutch early child language: an effect of input ? J Child Lang, 28, 629–660.

Yang, C. (2002). Universal Grammar, statistics or both? *Trends in Cognitive Sciences*, 8(10): 451-456.