Speech Perception and Language Acquisition in the First Year of Life

Judit Gervain[1] and Jacques Mehler[2]

[1]Department of Psychology, University of British Columbia, Vancouver British Columbia, V6T 1Z4, Canada

[2]Neuroscience Sector, Scuola Internazionale Superiore di Studi Avanzati, Trieste 31014, Italy; email: mehler@sissa.it

Key Words

infancy, learning mechanisms, phonological bootstrapping, evolution of language

Abstract

During the first year of life, infants pass important milestones in language development. We review some of the experimental evidence concerning these milestones in the domains of speech perception, phonological development, word learning, morphosyntactic acquisition, and bilingualism, emphasizing their interactions. We discuss them in the context of their biological underpinnings, introducing the most recent advances not only in language development, but also in neighboring areas such as genetics and the comparative research on animal communication systems. We argue for a theory of language acquisition that integrates behavioral, cognitive, neural, and evolutionary considerations and proposes to unify previously opposing theoretical stances, such as statistical learning, rule-based nativist accounts, and perceptual learning theories.

Contents

INTRODUCTION

The emergence of language has intrigued scientists and the general public alike, but it was only in the second half of the twentieth century that a systematic empirical investigation of language acquisition began. This work was greatly inspired by the suggestion that the environment is mainly a trigger rather than a tutor for language acquisition, at least during the first years of life (Chomsky 1959). Consequently, to explain the uniquely human capacity of language, scholars proposed innate acquisition mechanisms, specific to language (Chomsky 1959). A few years later, research into the biological foundations of language was expanded, giving a better grasp of the innate dispositions

for language acquisition (Lenneberg 1967). By contrast, other researchers suggested that classical learning mechanisms, ones that humans share with other animals, may be sufficient to acquire language (Elman 1996, Tomasello 2000). Under this view, the human specificity of language arises from quantitative rather than qualitative differences between the species.

Some of these theoretical questions may be resolved by studying preverbal infants, in particular newborns, as this allows us to determine how much of our language acquisition abilities are due to dispositions detectable much before the surroundings have shaped our cognitive apparatus. Therefore, our review mostly focuses on the development of language

and its underlying mechanisms during the first year of life. This choice is also justified by the growing body of research and recent advances in understanding how different mechanisms, such as statistical and distributional learning, rule extraction, as well as perceptual and memory constraints, work together during language development.

Our review discusses landmarks in language acquisition as well as their biological underpinnings. We focus on studies that connect brain, mind, and behavior. We believe that building bridges between these different levels is the way of the future and that the next decades will see the success of such integrative methodology and theory building.

In the review, we first describe the different theoretical approaches to language acquisition. We then review the increasingly important and fast-growing body of literature on the biological foundations of human language, focusing mostly on genetic and evolutionary aspects. Then we review the empirical evidence that has accumulated over the past decades in support of the theories and approaches introduced. We discuss the findings following the levels of organization in language from phonology through word segmentation and lexical acquisition to grammar. Finally, we consider some of the novel empirical findings that relate to the neural basis of language acquisition and processing in newborns and young infants. Building on these empirical findings, we argue for an integrative theory of language acquisition, proposing that rule learning, perceptual bootstrapping, and statistical learning all contribute to different levels of language acquisition, and that the most interesting objective is to understand their interactions and the division of labor among them.

THEORETICAL APPROACHES

Language acquisition came to the forefront of cognitive and developmental research when Noam Chomsky (1957, 1959) pointed out that acquiring language poses a serious learning problem. Infants never receive explicit information about the structure of the grammar that generated the utterances they are exposed to. In the absence of structural information, the finite data set that infants receive as input is compatible with an infinite number of underlying rules or grammars—a challenge to learning known in philosophy and mathematics as the induction problem.

The most important theoretical approaches to language acquisition in the past 50 years have investigated this logical problem, proposing solutions to it or denying its existence.

Nativist Approaches to Language Acquisition

Language cannot be learned exclusively from the input, yet young infants seem to acquire it with remarkable ease. Therefore, Chomsky (1959) argued that the acquisition process has to be guided by innate knowledge. This logical argument gave rise to a nativist theoretical approach to language acquisition as well as a large body of related empirical research (for a representative summary, see Guasti 2002). This view capitalizes on the observation that although they are superficially different from one another, languages of the world share a large number of structural characteristics; for example, they all use lexical categories like functors (small grammatical words, such as *he*, *it*, *on*, *of*, *this*) and content words (e.g., nouns and verbs that carry lexical meaning, such as *flower*, *table*, *run*, *sing*). Under the nativist view, the universal features of language design are part of our species' biological endowment and are encoded in the language faculty as innate principles. By contrast, aspects of language structure that vary (e.g., the relative order of verbs and objects or whether a language allows pronominal subjects to be dropped) are assumed to be encoded by parameters, i.e., mental switches that implement all the universal options [e.g., a verb-object (VO) order and an OV order; licensing pronoun-drop or not].

This account assumed that infants are able to detect and extract abstract regularities from the input. Indeed, Marcus et al. (1999) showed

that 7-month-old infants are able to learn abstract, algebraic generalizations. In their study, infants were familiarized with an artificial grammar encoding an identity-based regularity (e.g., ABB: *wo fe fe*). In the test phase, babies showed longer looking times for items that were inconsistent with the grammar of familiarization (e.g., ABA) than for items that were consistent with it, indicating that they extracted the underlying regularity.

Under the principles and parameters view, language acquisition is mediated by setting the parameters to the values that characterize the native language. For instance, an English-learning infant will have to set the word-order parameter to VO, e.g., *eat an apple*, and the pro-drop parameter to negative, e.g., *It is raining*, but not **Is raining*, while a Japanese infant will set both parameters to the opposite value, e.g., *ringo-wo taberu* 'apple.accusative eat' "eat an apple" and *futte iru* 'raining is' "(it) is raining." However, parameters are defined over abstract linguistic entities such as verbs, nouns, and pronouns, so the infant still faces the problem of linking these abstract mental representations to actual physical entities in the speech signal (Pinker 1984).

One solution proposed to the linking problem is the use of bootstrapping mechanisms. These are heuristic learning mechanisms that exploit the universal correlations that exist between perceptually available, surface characteristics of a language and its abstract morphosyntactic properties. Three types of surface cues have been proposed to act as triggers for bootstrapping.

One approach (e.g., Pinker 1984) suggests that the relevant cue is of semantic/conceptual nature. By understanding the general meaning of some simple sentences and by knowing the meaning of some words, typically nouns, the infant can construct syntactic trees, given configurational universals, such as the phrase structure suggested by generative grammar or other linguistic theories, which are believed to be part of the innate language faculty. From these trees, the child can derive the syntactic rules of her mother tongue, which in turn help her parse and understand more complex sentences.

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

It is important to note that these two approaches build on already acquired linguistic knowledge. But how are these initial pieces acquired? A third approach, the one we are exploring here, suggests that morphosyntactic properties are signaled by their acoustic/phonological correlates (Mehler et al. 2004; Morgan & Demuth 1996; Nespor et al. 1996, 2008). As Morgan & Demuth (1996, p. 2) put it: "[T]hese accounts propose that information available in speech may contain clues to certain fundamental syntactic distinctions [...]." This approach, unlike the other two, assumes no prior linguistic knowledge on the part of the learner and thus may explain the earliest acquisitions. Nouns and verbs, for instance, are abstract lexical categories. However, in English, nouns often bear stress on the first syllable (record N: /ˈrekɔ:(r)d/) and verbs on the last (record V: /rɪˈkɔ:(r)d/) (Cutler & Carter 1987, Davis & Kelly 1997). The stress pattern, then, can act as a cue to the two categories. Although this is specific to English, there seem to be phonological and prosodic cues that might signal syntactic properties universally (Mehler et al. 2004; Morgan & Demuth 1996; Nespor et al. 1996, 2008). An important focus of our review, therefore, is not only to characterize how infants perceive and learn about the acoustic, phonetic,

and phonological aspects of language, but also to explore how these might bootstrap the beginnings of morphosyntax during the first year of life.

Perceptual Primitives in Language Acquisition

How the acoustic and phonological aspects of speech are related to underlying structure has received increasing attention recently. According to a recent proposal by Endress et al. (2009), language might recruit previously existing perceptual mechanisms or "primitives" and use their outputs to feed abstract linguistic computations. In the perception and memory literature, for instance, it has long been known that sequence edges are particularly salient positions, facilitating perception, learning, and recall of elements in those positions (see Endress et al. 2009 for a summary). This, the authors argue, might be related to why languages show a universal preference for word-initial and word-final morphosyntactic processes as opposed to word-internal ones; e.g., prefixing and suffixing are common among languages, whereas infixing is very rare. Indeed, Endress et al. (2005) have recently demonstrated that adult learners perform well in an artificial grammar learning task if the regularity that they need to learn (identical adjacent repetition of syllables) is at the edge of a syllable sequence, but they fail if the same regularity appears sequence internally.

Similarly, as Endress et al. (2007) have demonstrated, identical repetitions are perceived automatically as salient Gestalts by adult learners in artificial grammar paradigms. When participants' task was to learn a sequence of three tones where the second and third tones were identical, they succeeded. But they failed when the tone sequences implemented an ordinal regularity, for example, a high tone followed by a low tone followed by a middle tone. Repetitions or identity appears to be a special input configuration that is more readily perceived than are other relations of the same mathematical complexity, for example, ordinal relations.

In the following sections, we review how some perceptual primitives, for example, the detection of repetitions (Endress et al. 2005, 2007; Gervain et al. 2008a), edge salience (Endress et al. 2005, 2007), or prosodic grouping principles (Nespor et al. 2008), might help bootstrap the acquisition of morphosyntactic structure.

Statistical Approaches to Language Acquisition

Although the above described nativist position has been very influential in the past 50 years, the long tradition of empiricist approaches to language acquisition has re-emerged in the past two decades. These empiricist positions take different forms, from statistical learning approaches to connectionism (Elman et al. 1996); what they share, though, is a belief that no innate language-specific knowledge is required to explain language acquisition. Rather, language development is a piecemeal learning process that relies on general-purpose mechanisms, typically statistical in nature, shared by most perceptual and cognitive domains. No innate mental contents specific to language such as lexical categories, principles, or parameters are assumed.

These statistical learning approaches gained new momentum in the language-acquisition literature when Saffran et al. (1996) demonstrated that very young infants are able to use statistical information contained in speech and to then use such information to segment continuous speech into its constituent words. These initial results have given rise to a large body of research, partly reviewed in The Word Segmentation Problem section below, investigating the role, scope, and limitations of statistical learning in language acquisition.

These statistical accounts have also been combined with social learning theory. In Tomasello's (2000) account, infants begin by learning frequently occurring sequences in the input (e.g., *Where is the toy? Where is the cup? This is a ball. This is a dog.*). As a second step, infants discover similarities among these memorized sequences and extract semiabstract

constructions or templates with a memorized component and one variable element (*Where is the _? This is a _.*). In these templates, the variable elements are not variables in a mathematical sense, as their scope might be limited to an arbitrary set of elements, for example, the members of the family, animals, or cars. Abstract, adult-like linguistic knowledge is believed to emerge only later, as young children generalize further, using the semiabstract templates. In Tomasello's (2000) view, infants and young children are aided by their social learning abilities during the stepwise abstraction process. They understand and construct the meaning of utterances not solely on the basis of the semantics of the linguistic constituents in the utterances addressed to them, but also by inferring the possible meaning from the speaker's intention, which even very young infants have been shown to have access to (Csibra & Gergely 2009, Gergely & Csibra 2003, Onishi & Baillargeon 2005).

Our review takes an integrative stance, emphasizing that innate language-specific, perceptual, and statistical mechanisms are all necessary for language acquisition. What needs to be explored is their respective roles and the interactions between them.

EVOLUTIONARY ORIGINS AND BIOLOGICAL FOUNDATIONS: APES, BIRDS, AND HUMANS

The nativist position on language acquisition grounded language in human biology. The initial investigations focused on the neurobiology of language, citing critical period effects, language acquisition in congenitally blind and deaf children, neurally based language pathologies, etc. (see Lenneberg 1967 for a classical formulation). More recently, in an attempt to investigate the most fundamental questions about language, numerous papers have explored its evolution. In parallel, studies of nonhuman animals are proceeding in the hope of determining whether human abilities have arisen in the human mind as a patchwork of different precursor systems that were present in ancestral species.

This line of research is of particular relevance for language acquisition because it raises convergent theoretical questions about innate, genetically endowed language abilities. If a predisposition for language is innate in humans, it became part of our genetic heritage during evolution. Therefore, research into nonhuman species' cognitive and communicative abilities complements studies of early infancy. Such comparative research also sheds light on the issue of language specificity. If humans and nonhuman animals share cognitive and/or learning abilities, these cannot be language specific since only our species has language. However, they may have been precursors bringing humans closer to language.

Research comparing human (infant) language acquisition and nonhuman cognitive, perceptual, and learning abilities usually takes one of two routes. Traditionally, humans' capabilities were compared to those of their closest evolutionary relatives, primates. Indeed, comparative studies between infants and primates have shown that the latter are also capable of statistical learning (Newport et al. 2004), language discrimination on the basis of rhythm (Ramus et al. 2000), and categorical phoneme perception (Morse et al. 1987), among other abilities. More recently, birdsong has been explored as a possible analogy for human language. This may, at first, appear surprising, since songbirds are not closely related to humans. However, vocal communication, like human language, plays an important role in songbirds' cognitive as well as social development, which is not the case for nonhuman primates. Songbirds' sophisticated vocalization system thus allows us to investigate not only learning and cognitive abilities underlying language as an abstract system, but also the mechanisms involved in vocalization, i.e., the relationship between perception and production. In addition, birdsong is highly complex, which allows a better comparison with human language than structurally simpler primate calls do. To quote Prather et al. (2009), "all songbirds studied to date [...] learn their song notes by imitation, a feature of human speech that is otherwise rare

among animals [. . .]. Swamp sparrows' songs comprise repeated groups of 2–5 'notes', which are composed of short pure-tonal frequency sweeps, with note categories differing primarily in duration, bandwidth and rate of change in frequency."

Below we show the relevance of birdsong as a comparative model of speech, if not necessarily of language. Investigating birdsong from this perspective gives us the opportunity to dissociate evolutionary ancestry from adaptive pressures. Phylogenetically different vocal communication systems might have developed similar mechanisms not because of common ancestry, but as a response to similar environmental and adaptive pressures. Comparing human language to birdsong makes it possible to explore the components of human language that are the result of selection and those that arose through hereditary endowment.

Genetic Studies of Speech Production and Language

Mutations in FOXP2 cause speech, morphological, and in all likelihood, other language disorders (Gopnik & Crago 1991, Haesler et al. 2007, Marler & Peters 1981). Haesler et al. (2007) began to study whether birds also possess behaviors and neural structures related to FOXP2 mutations after patients suffering from speech dyspraxia were found to have functional abnormalities related to high levels of FOXP2 in the striatum and basal ganglia. They reasoned that if birds also had problems related to elevated levels of FOXP2, then it would be possible to use birds as a model to understand whether the genetic underpinnings of speech were similar to those of birdsongs. The authors used zebra finches because they learn their songs by "imitating" adult tutors and because they change songs seasonally. Haesler et al. (2007) noticed that the expression of FOXP2 tends to increase in Area X when zebra finches learn to sing. The levels of FOXP2 decrease before the birds begin to learn their songs. The authors experimentally lowered the level

of FOXP2 in Area X during song learning and found that the experimental birds with decreased levels of FOXP2 sing in atypical ways as compared with controls. This study suggests that songbirds have mechanisms for learning their songs that are reminiscent of humans learning to speak and are susceptible to mutations in FOXP2. Since these findings, several other experiments have enriched our understanding of the expression of the genetic endowment and learning abilities (e.g., Miller et al. 2008).

Similarities Between Birdsong and Human Speech

The similarities of some mechanisms observed in songbirds and humans are indeed quite striking. Birdsong and human speech might use similar brain mechanisms: Auditory brain areas responsible for perception and motor areas responsible for production might be closely linked in both systems with single neurons responding to both perceived and produced vocalizations. For humans, the motor theory of speech, linking perception and production, was proposed decades ago (Liberman et al. 1967). More recently, Prather et al. (2008) identified similar mechanisms in swamp sparrows. The brain area HVC (high vocal center) of male swamp sparrows is engaged during song production, song perception, and learning of songs from tutors. Prather et al. (2008) investigated whether HVC neurons display both types of activity by recording from this area in freely behaving male swamp sparrows during presentation as well as production of songs. The authors found that some HVC neurons were active during singing *and* listening, which, as the authors demonstrated, was due to a motor estimation of auditory feedback. To confirm that this activity is indeed motor in nature and not simply due to auditory feedback as the bird perceives its own song, the authors played different distracting songs to birds while they were singing, so auditory feedback was disrupted. Increased neural activity was observed despite this manipulation. This, as the authors suggest, bears

resemblance to the motor theory of speech perception (Liberman et al. 1967) as well as to the mirror neuron system in the frontal cortex of monkeys (Gallese et al. 1996, Rizzolatti et al. 2001).

Birdsong has been suggested as a potential analog for speech and/or language due to its complex structure. In birdsong, just like in human language, the origin of this structural complexity, whether it is genetically determined or learned, is an exciting question. Feher et al. (2009) have asked whether species-typical songs can be created de novo in zebra finches, much like language can emerge in groups of linguistic isolates in the span of a few generations (Senghas et al. 2004). Feher et al. (2009) studied juvenile birds, raised in isolation. Songs that are usually observed in isolated (ISO) birds are less structured, noisier, and contain high-pitch upsweeps, making it possible to quantify the differences observed between the wild-type (WT) and ISO type of songs. Each juvenile bird was trained by a particular ISO tutor in a sound-proof cage. A number of isolated birds served as individual tutors to teach juveniles who had been deprived of prior exposure. Pupils of the first generation become tutors for other juvenile isolates, an operation that went on until the fourth generation was reached. Changes were observed in each successive training stage from the first to the fourth generation. The data show that the WT and ISO songs differ in their spectral features and duration of the acoustic state of songs, but across generations there is a progression from the ISO toward the WT song properties. The authors claim that "song culture is the result of an extended developmental process, a 'multigenerational' phenotype partly genetically encoded in a founding population and partly in environmental variables, but taking multiple generations to emerge." These findings bear strong resemblance to language emergence de novo in that more structured and more species-typical song and language emerge as a result of the acquisition/learning process, suggesting that impoverished input is sufficient to trigger the genetically encoded mechanisms responsible for song/speech.

The above reviewed evidence indicates that similarities between birdsong and speech exist at the level of neural mechanisms as well as in terms of the underlying genetic bases. But is birdsong a good model for the core property of human language, namely its structural complexity? Gardner et al. (2005) looked at canaries (*Serinus canaria*), which produce hierarchically organized songs. Songs consist of "syllables," which, when repeated, form a "phrase." Such phrases appear in young canaries after 60 days when they are raised typically, that is, in a population of singing adults. It is known that deafened juveniles produce the species-specific hierarchical organization, although syllables and phrases are impoverished. Gardner et al. (2005) exposed isolated juveniles to synthesized songs that were "ungrammatical" because they implemented a "random walk" through the syllable space. Initially, the production of the isolates seemed congruent with the random walk exposure. Upon transition to adulthood, however, normal syllables became recognizable and primitive phrasing started to emerge. At the end of the learning process, juveniles produced standard syllables, and species-typical phrasing was clearly noticeable. The authors concluded that "imitation and innate song constraints are separate processes that can be segregated in time: freedom in youth, rules in adulthood."

Gentner et al. (2000) and Prather et al. (2009) further examined song organization and perception in birds, focusing on categorical perception (for a discussion of categorical perception in humans, see The Early Sensitivity to Speech and Subsequent Phonological Development section). In the latter study, the authors systematically manipulated note duration, a learned aspect of swan sparrow song, and found that sensorimotor neurons showed a categorical response to gradually varying note duration. This neural response coincided with category boundaries observed behaviorally in the animals. Furthermore, sparrows coming from song dialects exhibiting different categorical boundaries responded according to the boundaries of their own species, indicating that boundaries were indeed learned.

In sum, it appears that birdsong and human speech are comparable in terms of their underlying neural mechanisms, the presence of innate guiding principles as well as some of their organizational properties. This, of course, does not imply that birdsong is equivalent to human language in terms of its productivity and structural complexity. Nor does it mean that songbirds' cognitive abilities are more similar to those of humans than are the cognitive abilities of primate species. Comparisons with birdsong provide us with an optimal testing ground to explore the genetic ancestry as well as the adaptive pressures that have shaped human language during the evolution of our species.

These similarities notwithstanding, human language appears to have a unique productivity and computational power not paralleled in any other species. Where do these features originate? After reviewing the abilities and mechanisms shared by humans and other animals, we turn to those that might be unique to our species.

Does a New Computational Component Cause the Emergence of Language in the Human Brain?

In an influential paper, Hauser et al. (2002) proposed that enquiries into language evolution should be incorporated into theories of language. They suggested that it may be convenient to distinguish between two aspects of the human language faculty: the language faculty in the broad sense (FLB) and the language faculty in the narrow sense (FLN). Their proposal is that the FLB is composed of various elements such as sensory motor systems, memory systems, social abilities, and so forth, whereas the FLN comprises a very limited number of computational components or a single computational component, which the authors view as quite likely to have been sufficient for the emergence of language. A similar conclusion has been drawn by other researchers with respect to mathematical abilities. "The human species is unique in its capacity to create revolutionary cultural inventions such as writing and mathematics, which dramatically enhance its native competence. From a neurobiological standpoint, such inventions are too recent for natural selection to have dedicated specific brain mechanisms to them. It has therefore been suggested that they co-opt or 'recycle' evolutionarily older circuits with a related function [...], thus enriching (without necessarily replacing) their domain of use" (Knops et al. 2009, p. 1538).

This way of presenting the theoretical framework proposes that many components (use of the vocal tract, categorical perception, etc.) are present in other animals. For a detailed discussion of which phonological abilities might be found in nonhuman species, see Yip (2006). The computational abilities required to acquire the syntax of the native language, by contrast, are unique to humans. Hauser et al. (2002) framed their paper as "a quest for the crucial evolutionary step that allowed our species to acquire the complex syntax of human languages."

In a follow-up experimental paper, Fitch & Hauser (2004) and Saffran et al. (2008) proposed that recursion, responsible for discrete infinity, might be the one and unique component of FLN. This proposal generated a great number of experiments and theoretical debates seeking to support or infirm the conjecture (Bahlman et al. 2006, Fitch et al. 2005, Friederici et al. 2006, Hauser et al. 2002, Hochmann et al. 2008, Pinker & Jackendoff 2005). Fitch & Hauser (2004) based their studies on the complexity of grammars that Chomsky (1957 and subsequent work) proposed. Chomsky made the claim that human languages are best characterized as context-free or phrase-structure grammars (PSG), not as computationally more limited finite-state grammars (FSG). Fitch & Hauser (2004) report an experiment investigating whether humans and monkeys are similar in their abilities to learn a FSG and a PSG from the simple presentation of items derived from the grammars. The authors used two artificial grammars. The FSG had items conforming to structure $(AB)^n$ with $n \leq 3$, the PSG to structure $A^n B^n$

with n ≤ 3. The authors habituated humans and cotton-top tamarin monkeys to either of these items. As and Bs were consonant-vowel syllables, with a female voice pronouncing the A syllables and a male the B syllables. In the test phase, humans had to rate new items as congruent or incongruent with the grammar they had learned, whereas monkeys were tested with a head-turn procedure to estimate whether the underlying grammar had been extracted. Humans behaved as if they had learned both grammars and monkeys as if they had the capacity to extract only the FSG grammar.

Later, Gentner et al. (2006) studied European starlings and challenged the notion that only humans can learn PSG. They used the same kinds of grammars as had Fitch & Hauser (2004), except that As and Bs corresponded to two specific categories of sounds these birds use. Before being tested, birds were trained with a protracted operant-conditioning schedule, a procedure that Fitch & Hauser (2004) did not use with the cotton-top tamarins. After this extended training phase, starlings learned the PSG.

Perruchet & Rey (2005) criticized Fitch & Hauser (2004) on different grounds, arguing that in Fitch & Hauser's (2004) study, humans did not actually need to establish nonadjacent dependencies to succeed and cannot therefore be assumed to have extracted the underlying structure of the $A^n B^n$ items. Indeed, the distributional properties and/or the rhythmic properties of Fitch & Hauser's (2004) material offer a better explanation of how humans processed the $A^n B^n$ items. Indeed, Hochmann et al. (2008) showed that human participants in the test did not dismiss $A^2 B^3$ or $A^3 B^2$ as incongruent with the grammar $A^n B^n$. Moreover, when interrogated at the end of the experiment, those few participants who did dismiss such items reported that they explicitly counted the number of As and Bs and only accepted sequences with equal numbers. Despite these empirical issues, the theoretical proposal made by Hauser et al. (2002) remains highly interesting and invites further research.

We follow this brief review of the evolutionary aspects of human language and animals' abilities with a detailed discussion of young infants' speech and language-processing capacities to provide an empirical basis for the evaluation of the theoretical and evolutionary claims introduced so far.

THE EARLY SENSITIVITY TO SPEECH AND SUBSEQUENT PHONOLOGICAL DEVELOPMENT

Newborn infants show surprising speech-processing abilities from birth. They prefer forward-going speech and primate vocalizations over acoustically matched nonspeech sounds or backward speech (Dehaene-Lambertz et al. 2002; Pena et al. 2003; Vouloumanos & Werker 2004, 2007), their mother's voice over other female voices (Mehler et al. 1978), and their native language over unfamiliar languages (Mehler et al. 1988, Moon et al. 1993). These early language discrimination abilities might represent some form of imprinting to the properties of the native language upon the first encounter immediately after birth, or alternatively the result of exposure to the maternal language in utero. Newborns can make most of the phonemic distinctions attested in the world's languages (Dehaene-Lambertz & Dehaene 1994, Eimas et al. 1971, Werker & Tees 1984b), and they are able to distinguish languages they have never heard before on the basis of their rhythmical characteristics (Mehler et al. 1988; Nazzi et al. 1998; Ramus et al. 1999, 2000). Newborns are also able to detect the acoustic cues that signal word boundaries (Christophe et al. 1994), discriminate words with different patterns of lexical stress (Sansavini et al. 1997), and distinguish function words (e.g., *it*, *this*, *in*, *of*, *these*, *some*) from content words (e.g., *baby*, *table*, *eat*, *slowly*, *happy*) on the basis of their different acoustic characteristics (Shi et al. 1999). These early, innate abilities lay the foundations for later language learning.

Acquisition of the Native Phonology

One of the most fundamental and at the same time most surprising perceptual abilities of newborns is that they are able to discriminate most sound contrasts used in the world's languages. In other words, they are born as "citizens of the world," ready to learn any natural language. Just like adults, newborns perceive these sounds categorically (Eimas et al. 1971, Liberman et al. 1957), perceiving acoustic variation from within a phoneme boundary as the same sound and the same acoustic variation spanning adult phoneme boundaries as being different sounds.

During the first year of life, as a result of exposure to the native language, this initial universal discrimination narrows down to the phonemes, that is, minimal meaningful differences (e.g., *pin* versus *bin*), of the native language (Werker & Tees 1984a). Discrimination of most nonnative contrasts is lost (Werker & Tees 1984a), whereas it is maintained or even enhanced for native contrasts (Kuhl et al. 2006). English, for instance, only has a dental /d/ sound, whereas Hindi discriminates between a retroflex /D/ and a dental /d/. Newborns and young infants born into English-speaking environments readily discriminate the Hindi sounds. But after eight months of exposure to English, where the two categories are not distinguished, English-learning infants start losing the discrimination (Werker & Tees 1984a). Indeed, English-speaking adults find it very hard to discriminate this contrast. Hindi infants and adults, as a result of exposure to Hindi, maintain it.

What learning mechanism might account for this learning-by-forgetting (Mehler 1974, Mehler & Dupoux 1994) or perceptual attunement (Scott et al. 2007) process? It has been suggested that native phonological categories might be established through a distributional learning mechanism (Maye et al. 2002). In a language like English, where there is only one /d/ sound, most actual realizations that infants encounter will cluster around a prototypical /d/ pronunciation, so the distribution of English /d/ sounds will have a mode around the most typical acoustic parameters for /d/. On the other hand, in Hindi, where there are two /d/ sounds, the same acoustic space will show a bimodal distribution, as there will be many instances around the typical /D/ sound as well as around the typical /d/ sound. As a result, in English, infants will be exposed to a unimodal distribution, and in Hindi, a bimodal one. It has been shown that infants are sensitive to this statistical distribution, and they create a single phoneme category when exposed to a unimodal distribution, whereas they establish two categories if the distribution in the input is bimodal (Maye et al. 2002). In their study, Maye and colleagues (2002) used the /da/-/ta/ continuum, where the two syllables are distinguished by the onset of voicing (voice onset time, or VOT). Since /d/ is a voiced consonant, in /da/, voicing starts at 0 msec, that is, immediately at the onset of the syllable, whereas in /ta/, the consonant is voiceless; thus, voicing starts only at the onset of the vowel. By delaying VOT incrementally, a continuum was created from /da/ with VOT at 0 msec through six syllables with VOT at 20 msec, 40 msec, etc., to /ta/ with VOT at 140 msec. One group of 6- to 8-month-old infants, the unimodal group, was exposed to a frequency distribution along this continuum where syllables in the middle (instances 4 and 5 with VOT 60 msec and 80 msec, respectively) had the highest frequency of occurrence. A second group, the bimodal group, was exposed to a distribution where tokens closer to the end points (with VOT 20 msec and 120 msec) were the most frequent ones. When tested on the discrimination of the end points of the continuum, /da/ and /ta/, the bimodal group showed better discrimination than the unimodal group (Maye et al. 2002).

These results suggest that infants have the ability to track the frequency of sound tokens in the input and might use this information to tune into native phonemic categories (Best & McRoberts 2003, Kuhl 2004, Maye et al. 2002).

The Early Sensitivity to Rhythm and Its Potential Bootstrapping Role

The previous sections have illustrated the challenge of acquiring one's native language. However, some infants successfully acquire not only one, but two or more languages at the same time. How do these infants discriminate between their languages?

Linguists have long recognized that languages differ perceptibly in their sound patterns and, in particular, in their rhythm (Abercrombie 1967, James 1940, Ladefoged 1993, Pike 1945). Initially, these differences were described as categorical and were derived from the isochrony principle, that is, as a function of the linguistic unit that has a constant duration in a given language. According to this view, languages fall into one of three rhythmic classes. In stress-timed languages such as English, Dutch, or Polish, the isochronous unit is the time between two subsequent stressed syllables. For example, in the sentence *Pronunciation is important in English*, the duration of time between the stressed syllables (in bold) is roughly the same. In syllable-timed languages, such as Spanish or Italian, the unit of isochrony is the syllable, that is, syllables are roughly of equal duration. For instance, in *tavolo* 'table' (Italian), no vowel is reduced, so all syllables are of the same length. In mora-timed languages, such as Japanese or Tamil, the isochronous unit is the mora. The mora is the measure of syllable weight [light/short syllables such as *a* (the indefinite article) consist of one mora; heavy/long syllables such as *see* consist of two morae].

These differences in rhythm are intuitive and easy to perceive for adults. If infants have the same sensitivity to linguistic rhythm, it might help them discriminate their languages, at least when those are from different rhythmical classes. Such an early sensitivity was indeed observed by Mehler et al. (1988), who showed that newborns were able to discriminate their future native language from a rhythmically different language, even if both were low-pass filtered, suppressing phoneme identity. This initial finding, suggesting that language discrimination relies upon suprasegmental, rhythmical cues, was extended by Nazzi et al. (1998), showing that rhythmical differences were sufficient for discrimination; familiarity with the languages was not necessary. These authors found that French newborns readily discriminated between low-pass filtered utterances in English and Japanese, two languages they had never heard before.

These results established that rhythm might serve as an initial cue to language discrimination. However, the exact acoustic features corresponding to the subjective percept of rhythm were still unknown. The isochrony principle proved incorrect, as empirical investigations obtained no isochrony for the relevant units (Dauer 1983), and several languages were found that showed characteristics of both stress-timed and syllable-timed rhythm (Nespor 1990). Rhythmicity thus appeared to be a gradient rather than a categorical property (Nespor 1990). Building on these observations, Ramus et al. (1999) proposed an operational definition for rhythm and rhythmical classification as a function of three acoustic parameters: (*a*) %V, the proportion of vowels/vocalic space relative to the total length of an utterance, (*b*) ΔV, the variability in the length of vocalic spaces, and (*c*) ΔC, the variability in the length of consonant clusters. The authors measured these parameters in naturalistic recordings of speech in eight languages (e.g., English, Dutch, French, Italian, Japanese) and found that languages clustered into groups similar to the traditional rhythmical classes when plotted in two-dimensional spaces defined by any two of the three acoustic parameters. This definition recreated the traditional classification and accounted for languages previously found to be ambiguous (Nespor 1990) with respect to classification or currently undergoing change, because continuous rather than categorical measures were used. It is important to note that work by Grabe & Low (2002), also using a computational definition of rhythm, failed to recreate the traditional rhythmic classes. However, as subsequent work by Ramus (2002) suggests, there were important methodological

differences between Ramus et al.'s (1999) and Grabe & Low's (2002) studies, which might account for the different findings. Grabe & Low (2002) analyzed speech from one speaker per language, whereas Ramus et al. (1999) recorded four speakers for each language, thus obtaining a measure that matched the general pattern of languages more closely than did the idiosyncrasies of individual speakers.

The classification in terms of %V, ΔV, and ΔC suggested that it wasn't specific segmental identity that defined rhythm, but rather the relative length and variability of vocalic and consonantal spaces. Ramus & Mehler (1999) and Ramus et al. (1999) tested this prediction in a series of experiments in which they replaced individual vowels by /a/ and individual consonants by /s/. Utterances resynthesized this way suppressed phonemic and consequently lexical identity, but preserved the proportion of vowels and consonants in the signal. Adults as well as newborns were able to discriminate utterances from two rhythmically different languages when this resynthesis was applied. However, they failed when both vowels and consonants were transformed into /a/, suppressing the difference between them. These results clearly established that the three parameters relating to the ratio of vowels and consonants in the speech signal were necessary and sufficient acoustic cues for rhythm-based language discrimination at birth. The discrimination of rhythmically similar languages emerges at around age 4 months; it has been hypothesized to rely on more subtle cues, such as phoneme identity or phonotactics (Bosch & Sebastián-Gallés 2001, Ramon-Casas et al. 2009).

In addition to language discrimination, linguistic rhythm might also serve as a bootstrapping cue for morphosyntax. Languages belonging to different rhythmic classes also show different morphosyntactic properties. For instance, mora-timed languages, that is, languages with a high value for %V, such as Japanese, tend to have simple syllabic structure, agglutinating morphology, and object-verb (OV) word order, whereas languages with

lower %V values, such as English or Polish, typically have complex syllable structure, inflecting morphology, and VO word order (Fenk-Oczlon & Fenk 2005). Given these correlations, Mehler et al. (2004) have proposed that rhythm might act as a bootstrap for general morphosyntactic type. The proposal hasn't been tested empirically, but it is of potential importance because it links a robust acoustic cue, detected even by neonates, to the most general and most abstract morphosyntactic properties.

THE WORD SEGMENTATION PROBLEM: LEARNING MECHANISMS AND PERCEPTUAL PRIMITIVES

Parallel to the task of breaking the syntactic code of their native language, infants also need to start building a lexicon. According to an increasingly widespread view (see Swingley 2009 for a review), lexical acquisition starts as early as the second half of the first year of life, when infants begin to segment potential word forms out of the continuous speech stream they hear. These forms are believed not yet to be reliably associated with meaning; nevertheless, they play a significant role not only in building the lexicon, but also in morphosyntactic acquisition. In other words, lexical acquisition starts much before infants utter their first words.

Learning word forms is a challenging task since speech is continuous: Most word boundaries are not marked by pauses, and words typically do not occur in isolation. Yet the sensitivity to potential word forms appears as early as birth. Newborns are able to discriminate identical phoneme sequences that only differ in that some span a word boundary, whereas others don't (e.g., *panorama typique* versus *mathématicien*, respectively; Christophe et al. 1994). This result provides a good example of infants' early sensitivity to perceptual Gestalts like edges and to prosodic structure in general. In addition, newborns are also able to discriminate word forms with different patterns of lexical stress (Sansavini et al. 1997).

These early sensitivities notwithstanding, extracting and storing a relatively large number of word forms from speech starts only at about age 6 to 8 months. Several mechanisms have been proposed to account for this feat. Statistical learning has been proposed as a general-purpose, potentially universal mechanism, which might be operational early on, whereas language-specific mechanisms, which require some familiarity with the native language, such as tracking allophonic variation, phonotactics, or stress patterns, are suggested to emerge somewhat later (Swingley 2005).

Statistically Based Word Segmentation

Proponents of structural linguistics (Harris 1955) and information theory (Shannon 1948) have long recognized that the statistical information encoded in language provides cues to its constituent units (e.g., morphemes and words) and structural patterns. Some words are much more frequent, that is, more probable, than others in absolute terms (e.g., *this*, *it*, *in*, *are*, *dog*, *time*) or in a given context (e.g., *chips* after *fish and...*; *do* or *is* at the beginning of a sentence).

Building on these observations, Hayes & Clark (1970) tested whether adult participants can use statistical information to extract words from a continuous stream of sine-wave speech analogs and found successful segmentation. Later, Saffran et al. (1996) showed that 8-month-old infants could use statistical information, more specifically transition probabilities (TPs; i.e., the probability with which one syllable predicts the next or the previous one), to segment a continuous stream of syllables, where syllables within a word predicted one another with a probability of 1.00, while syllables spanning word boundaries had TPs of 0.33. Infants could use dips in TPs to identify word boundaries.

Statistical learning has been shown to be a robust, domain-general, age-independent, and not specifically human ability. It operates over speech sounds as well as tones (Kudo et al. 2006) and visual stimuli (Fiser & Aslin 2002a,b). It is performed by newborns (Teinonen et al.

2009), infants at 8 and 13 months (Saffran et al. 1996), and adults (Pena et al. 2002). Moreover, nonhuman species, such as tamarin monkeys (Hauser et al. 2001) and rats (Toro & Trobalon 2005), are also able to learn statistical information.

Perceptual and Linguistic Constraints on Statistical Learning

Saffran et al.'s (1996) results shed new light on the well-known fact that humans are powerful statistical learners. But how is statistical learning used in language acquisition? A recent set of studies suggests that statistics are not used across the board for learning language. Rather, they are recruited for specific learning tasks—in particular, word segmentation and lexical acquisition—triggered by cues in the speech signal, and their application is limited by linguistic constraints.

Inspired by the fact that both morphology and syntax make use of constructions with distant dependencies, Pena et al. (2002), Newport & Aslin (2004), and Newport et al. (2004) asked the question whether transition probabilities between nonadjacent items can be learned. Pena et al. (2002) found that adults readily segmented out trisyllabic words from an artificial language when they were defined by high TPs between the first and the last syllables (A X C). However, subjects failed to generalize the pattern to novel X items unless (subliminal) segmentation cues were inserted into the stream to facilitate the original segmentation task, allowing participants to better process the regularity (Pena et al. 2002). These results suggest that cues in the signal, for example, pauses, act as triggers for different processing mechanisms, for example, statistics versus rule generalization.

A second and related issue that arises is the nature of the units or representations to which statistical computations apply. Bonatti et al. (2005) observed that adults readily segment over nonadjacent consonants, but not over nonadjacent vowels. This finding was further confirmed by Toro et al. (2008), who devised

a series of artificial grammar experiments to show that consonants and vowels serve as preferential input to different kinds of learning mechanisms. They found that participants performed well when their task was to do statistical computations over consonants or rule-learning over vowels (the rule to be learned was a repetition-based generalization). But their performance dropped to chance in the opposite case, i.e., statistical computations over vowels and rule-learning over consonants. Taken together, these studies indicate that not all linguistic representations are equally suitable for statistical learning. Consonants seem to be the primary target, while vowels are preferentially recruited for rule learning.[1] These findings converge with certain observations in linguistics (Nespor et al. 2003) suggesting that consonants and vowels have different linguistic functions. Consonants are believed to be responsible for encoding the lexicon; e.g., consonantal stems carry the semantic contents of lexical items in Semitic languages. By contrast, vowels are claimed to signal morphological form and syntactic function, e.g., Ablaut phenomena in Germanic languages, *sing*, *sang*, *sung*. These studies provide further evidence that statistical computations are selectively triggered and constrained by cues in the input, and their primary function is lexical segmentation.

However, the use of statistics for segmentation and word-form learning might not be universal. In some languages, such as Chinese or infant-directed English, most words are monosyllabic, rendering statistical computations vacuous (Yang 2004, Yang & Gambell 2004). Morphologically complex languages, such as Hungarian (*ház-a-i-nk-ból*

'house.possessive.plural.1stpl.from' "from our houses") and Turkish, might pose the opposite problem, as it is not clear what unit would be segmented out: complex word forms or individual stems and suffixes.

Taken together, these studies indicate that statistical segmentation alone is not sufficient to solve the task of extracting word forms from continuous speech. Other cues, taking into account the morphophonological properties of individual languages, are needed to complement statistical computations.

Language-Specific Cues to Segmentation

Although words are not separated by clear pauses in continuous speech, there are some acoustic and phonological features that correlate reliably enough with word boundaries to allow successful segmentation in most cases. At least three such cues have been identified in the literature, mostly on the basis of English: word-level stress patterns, phonotactic regularities, and allophonic variation.

Many languages assign word-level stress to a specific position within words; for example, Hungarian, has strictly word-initial stress. But even in languages where stress is not fixed but is lexically determined for each word, there are predominant patterns that can serve as heuristic cues. In English, word-level stress is lexically determined, but most bisyllabic nouns follow a strong-weak, that is, trochaic pattern (e.g., *doctor*, *infant*) Thus, segmenting speech at strong syllables is a potentially useful heuristic known as the metrical segmentation strategy (Cutler 1994, Cutler & Carter 1987). Indeed, Jusczyk et al. (1999) found that 7.5-month-old English-exposed infants show a trochaic bias, treating heavy syllables as word-initial (***doc**tor*, ***can**dle*). Importantly, the bias required words to be multisyllabic. Heavy monosyllables (*dock*, *can*) were not recognized (Jusczyk et al. 1999), but trisyllabic words with initial stress (strong-weak-weak) were treated as familiar, whereas weak-strong-weak and weak-weak-strong patterns were not (Curtin et al.

---

[1] It needs to be noted that Newport & Aslin (2004) found successful statistical segmentation for vowels as well as consonants. However, they used an artificial speech stream that allowed immediate repetitions of the same word frame, making the statistical patterns highly salient, whereas Bonatti et al.'s (2005) and Toro et al.'s (2008) stream had no immediate repetitions. It seems, then, that vowels might also be used for statistical computations under special conditions, such as the informationally highly redundant stream used by Newport & Aslin (2004).

2001). Importantly, the metrical segmentation strategy is a heuristic tool, since some English bisyllables are not trochaic, but iambic (e.g., *gui'tar*). In these cases, the strategy predicts initial missegmentation. This was confirmed empirically: 7.5-month-olds who readily recognize trochaic words in continuous passages failed to show similar recognition for iambs (Jusczyk et al. 1999).

Legal and illegal phoneme distributions, that is, phonotactics, also provide information about word boundaries. In English, the sequence /br/ is frequent word initially, but it is rare word internally. Therefore, it is a good candidate for a potential word onset. Conversely, words frequently end in /nt/, which is therefore a possible cue to the end of words. In a task where infants were exposed to CVCCVC (C, consonant; V, vowel) sequences with word-internally frequent or infrequent CC clusters, they segmented the sequences into two words in the latter case, but not in the former case (Mattys et al. 1999; Mattys & Jusczyk 2001a,b).

Variation in the realization of phonemes, known as allophony, can also indicate word boundaries. In English, for instance, aspirated stop consonants appear at the onsets of stressed syllables, whereas their unaspirated allophones appear elsewhere (Church 1987). At 9 months of age, infants are able to posit word boundaries based on allophonic (e.g., night rates versus nitrates) and distributional cues, and at 10.5 months, allophonic cues alone are sufficient for successful segmentation (Jusczyk et al. 1999).

The Interaction of Statistical and Language-Specific Cues

The above cues are mostly heuristic in nature and might lead to missegmentation in less frequent or atypical cases. Such missegmentations can be induced in experimental conditions (Jusczyk et al. 1999) and can also be observed in young children's spontaneous production (Slobin 1997). However, infants acquire the majority of the word forms they know without error. This implies that they are using more

than just one cue at a time, since converging cues yield more accurate segmentation.

Several studies have shown that young infants are indeed capable of using different cues simultaneously. When stress and phonotactic cues provide conflicting information about word boundaries, 9-month-old infants prefer to rely on stress cues (Mattys et al. 1999; Mattys & Jusczyk 2001a,b). When stress and statistical information are contrasted, 6-month-olds follow the statistical information (Saffran & Thiessen 2003), whereas 8-month-olds use stress cues (Johnson & Jusczyk 2001). This shift indicates a move from universal to more language-specific strategies as infants gain increasing familiarity with their native language.

Artificial grammar learning work with adults also indicates that statistical information and prosody are both computed in segmentation tasks, and prosody is typically used to constrain statistics in linguistically meaningful ways, as discussed above. If, for instance, the continuous speech stream is not monotonous as used in Saffran et al.'s (1996) original work, but has utterance-like intonational contours overlaid on it, then participants readily segment statistically coherent words inside prosodic contours, but not spanning two contours (Shukla et al. 2007). Similarly, while participants erroneously recognize "phantom words" in artificial speech streams, that is, words that never occurred in the stream, but their pair-wise syllable transitions have high probabilities (e.g., *fekula* was never heard, but *fe-ku* and *ku-la* appeared in the stream with high TPs), this false recognition can be suppressed if the stream contains prosodic cues to word boundaries, such as pauses or word-final lengthening (Endress & Mehler 2009).

Early Form-Meaning Associations

As suggested above, infants start learning words as early as age 6 to 8 months by extracting potential word forms from the input using statistical as well as phonological cues (see Swingley 2009 for a review). In order to develop a lexicon, they also need to start matching these word

forms with possible meanings. Learning the full meaning of words, especially abstract words or grammatical functors, requires advanced abilities, such as categorization, understanding referentiality, and solving the induction problem for meaning (Nazzi & Bertoncini 2003, Quine 1960, Waxman & Gelman 2009). We do not discuss these complex and advanced forms of word learning here. We only review the earliest stages of lexical acquisition, when a linguistic label gets associated with a perceptually available, concrete object.

These early associations were investigated by Stager & Werker (1997), who showed that infants use their phonological knowledge and representations differently at different stages of the word-learning process. At 8 months, before word learning en masse begins, infants readily discriminate a minimal pair of word forms, *bih* and *dih*, and they are also able to associate them with two different objects. At 14 months, which is the beginning of the word-learning stage, infants succeed in the simple phonetic discrimination task, but fail to distinguish the two words when they are used in a labeling context, that is, associated with two distinct objects. They succeed, however, even in this context if the words are very distinct, for example, *lif* and *neem*. At 17 months, when word learning is in full swing, infants succeed again in both tasks. The authors accounted for these results by arguing that phonological knowledge is recruited for word learning in different ways at different developmental stages. When starting to associate word forms with meanings, infants need to pay attention to the details of both and establish an association between them. At this early stage, infants might not attribute more importance to the minimal phonemic difference between two words than to other properties of the words, such as the speaker's gender. Given the high cognitive demands of the association task, a minimal phonemic difference might go unnoticed. At later stages, when infants become experienced word learners, label-object associations become less taxing for the cognitive systems; thus, even minor differences can be more readily utilized.

Confirmation for the cognitive load hypothesis comes from recent studies that found successful associations in 14-month-olds with minimally different labels when the cognitive load of the task was reduced, e.g., by using words known to the infants (Fennell & Werker 2003), by prefamiliarizing them with the objects (Fennell & Werker 2004), by giving them a visual choice between two objects in a test (Yoshida et al. 2009), or by making the acoustic difference between words more salient (Curtin et al. 2009) or more relevant for the task (Thiessen 2007).

BROAD LEXICAL CATEGORIES: FUNCTORS AND CONTENT WORDS

Words in the lexicon are organized into hierarchical categories. The most general and cross-linguistically universal divide is the one between closed-class functors (free or bound), such as articles, pronouns, and pre- or postpositions, and open-class content words, such as nouns, verbs, and adjectives. The most important difference between these two broad categories is functional: Functors signal morphosyntactic structure (e.g., plurality, tense, and argument structure), whereas content words carry lexical meaning. In addition, there are a number of statistical and acoustic/phonological differences between them. Functors have very high token frequencies. In corpora, they often account for 30% to 50% of the whole input (Gervain et al. 2008b, Kučera & Francis 1967). Content words typically have much lower token frequencies. By contrast, they are acoustically more salient, as they carry stress, consist of multiple syllables, and have at least one nonreduced vowel (Morgan et al. 1996).

It is well known that young children often omit functors in their early productions (Guasti 2002), which raised the question of whether they are able to perceive and represent functors at all. An early study (Shipley et al. 1969) showed that children whose linguistic production was at the telegraphic phase (i.e., contained no function words) nevertheless understood

instructions better if the instructions themselves were not telegraphic, but contained function words as well. Later, Gerken et al. (1990) established that the omission of functors in early production stems from a limitation on production and not on perception or encoding. In a series of imitation experiments with 2- to 3-year-old children, they found that children tend to omit weak, unstressed monosyllabic morphemes, typically functors, but not strong, stressed ones, typically content words, even if both are nonsense non-English words. Also, they imitate nonexisting content words with greater ease if they appear in the environment of real English function words as opposed to environments of nonsense function words. Moreover, children make a distinction between those nonsense functors that follow the usual consonant patterns of English functors and those that do not. Taken together, these results indicate that even though young children produce few functors, they build fairly detailed representations of them, which they can use in segmenting and labeling the incoming speech stream. In a later experiment, Gerken & McIntosh (1993) obtained similar results for sentence comprehension.

The above experiments were carried out with children who already have substantial knowledge of the grammar of their native language. But segmentation and labeling cues are most relevant at the beginning of acquisition to break up the input. Indeed, Shi et al. (1999) asked whether newborns are able to distinguish functors and content words on the basis of the phonological differences between them. Their findings indicate that newborn infants of both English-speaking and non-English-speaking mothers are able to categorically discriminate between English function and content words presented in isolation. By 6 months of age, infants start to show a preference for content words (Shi & Werker 2001), and by 11 months, they are also able to represent frequent functors in some phonological detail (Shi et al. 2006). They are also able to use functors, frequent and infrequent ones alike, to segment out a following content word (Shi et al. 2006).

Höhle & Weissenborn (2003) obtained similar results, showing functor versus content word discrimination in 7- to 9-month-old German infants exposed to continuous speech.

On the basis of the findings described above, it is not unreasonable to assume that the function word versus content word distinction is available to infants very early on, and although functors might not frequently appear in infants' earliest productions, they might be among their earliest word form representations, serving to bootstrap the early content words categories, e.g., nouns and verbs. Borrowed from the structuralist-generativist linguistic tradition, the idea that functors are fundamental for the categorization of content words has recently gained empirical support from corpus studies (Mintz 2002, Redington et al. 1998).

WORD ORDER AND OTHER TYPOLOGICAL DIFFERENCES

The acquisition and production of the first words at around the age of one year mark an important milestone in young infants' language development. Multiword utterances appear much later, after the second birthday. However, the acquisition of the most basic syntactic properties of the native language, such as word order, might actually start much earlier, during the first year of life, in parallel with and possibly in relation to early speech perception and word-learning abilities. Indeed, Brown (1973) has shown that infants get basic word order right from their first productions, which suggests that word order is a property that they have acquired prior to the production of multiword utterances.

How do infants acquire word order so early? According to the lexicalist position (Tomasello 2000), infants and young children initially do not represent word order in an abstract form. Rather, they learn relatively fixed constructions, often specific to individual lexical items, usually individual verbs (for example, *eat* is preceded by a noun phrase, the *eater*; and is followed by a noun phrase, the *eatee*). The generativist account, by contrast, assumes that even young

learners have general and abstract word-order representations encoding the relative order of the phrasal head and its complements and specifiers. For example, in a language with a head-complement, these technical terms are always spelled with capital initials; in linguistics, it would be better to follow this conventional order: objects follow verbs, nouns follow prepositions, etc. (e.g., *eat an apple*; *on the table*). One way to differentiate between these two accounts is to show that infants have some rudimentary representation of word order prior to the acquisition of a sizeable lexicon.

Recent results suggest that such a prelexical word-order representation might be created early on using frequency as a bootstrapping cue (Gervain et al. 2008b). As discussed above, functors are more frequent than content words. In addition, their position relative to utterance boundaries correlates with the general word order of languages (Gervain et al. 2008b, Morgan et al. 1996). In Italian, for instance, the general word order is VO; therefore, functors that head a phrase appear phrase initially (for example, prepositions: *sul tavolo* on-the table 'on the table'). By contrast, in Japanese, functors heading phrases are final (for example, postpositions: *Tokyo ni* Tokyo to 'to Tokyo'). In infant-directed speech corpora in these two languages, the distribution of frequent words, that is, functors, was exactly the opposite. In Italian, most two-word phrases at utterance boundaries started with a frequent word, that is, functor, whereas in Japanese, most of these phrases ended in a frequent word. Importantly, 8-month-old infants appear to be sensitive to these distributional differences. When exposed to a structurally ambiguous artificial speech stream in which frequent and infrequent nonwords alternated and the beginning and the end of the stream was ramped in amplitude to mask phase information, Japanese infants preferred to parse the stream into frequent-final units, whereas Italian infants showed longer looking times for frequent-initial test items (Gervain et al. 2008b). This suggests that prelexical infants show a rudimentary initial representation of word order, at least in terms of the relative

positions of frequent and infrequent words, that is, typically functors and content words. This finding has been confirmed by recent results (J. Hochmann, A. Endress, and J. Mehler, manuscript under review) suggesting that infants do indeed treat frequent words as functors and infrequent ones as content words. When infants were given the choice to pair either the frequent words or the infrequent words with objects, they chose the infrequent ones as possible labels for naming objects (J. Hochmann, A. Endress, and J. Mehler, manuscript under review).

However, unlike Italian and Japanese, some languages do not show a consistent word-order pattern. German, for example, uses both OV and VO orders within the verb phrase, depending on the syntactic context. Also, some infants grow up with an OV and a VO language simultaneously (for example, Japanese and English). In these cases, frequency alone does not provide enough information about word order, since both frequent-initial and frequent-final phrases occur in the input. This implies that further cues are necessary to bootstrap word order. One cue that has been suggested in the literature is prosody. Nespor et al. (2008) found that the location and the acoustic realization of prosodic prominence correlate with word order both across and within languages. Thus, in OV languages such as Turkish and in phrases with OV order within mixed languages such as German, prominence within prosodic phrases is initial, and it is implemented as a pitch contrast (high-low), whereas in VO languages such as Italian or French as well as in the VO phrases of mixed languages, a durational contrast is utilized, and prominence is final (short-long). If infants can use this prosodic cue in conjunction with frequency, then a more precise and fine-grained representation of word order can be acquired, even in cases where the two word orders, OV and VO, occur within a single language.

It has been argued that this grouping, that is, prominence-initial for pitch or intensity contrasts and prominence-final for durational contrasts, is an auditory bias that applies to speech and nonspeech stimuli alike (the

iambic-trochaic law; Hayes 1995). More recently, some data have been reported suggesting that the grouping principle might emerge as a result of language experience (Iversen et al. 2008; K.A. Yoshida, J.R. Iversen, A.D. Patel, R. Mazuka, H. Nito, J. Gervain, and F. Werker, manuscript under revision). However, these results are not conclusive, as other studies have found no language-related differences (R. Bion, S. Benavides, and M. Nespor, manuscript under review; Hay & Diehl 2007). Irrespective of whether this bias is independent of language experience or a result of it, infants might use it as a cue to word order at a very early age.

The hypothesis that even prelexical infants might possess some simple word order representations, possibly bootstrapped by frequency and prosody, received independent confirmation from studies using naturalistic stimuli in German. Weissenborn et al. (1996) found that German infants were sensitive to word order violations in German subordinate clauses.

THE NEURAL CORRELATES OF LANGUAGE IN YOUNG INFANTS

With the advancement of brain imaging techniques, it has become increasingly possible to pursue the original agenda of the research on the biological foundations of language with infant populations. Researchers have started charting the brain areas and circuits dedicated to language and speech perception in newborns and young infants.

One of the most important findings of this increasing body of research is that the newborn and infant brain shows a functional organization for language processing that is similar to that of the adult brain (Dehaene-Lambertz et al. 2002, 2008; Gervain et al. 2008a; Pena et al. 2003; Taga & Asakawa 2007). This organization appears to be at least partly under genetic control and develops even without experience with language (e.g., in congenitally deaf individuals; Dehaene-Lambertz et al. 2008).

More specifically, it has been observed that 3-month-old infants as well as newborns show a left-hemisphere advantage when listening to speech as compared with reversed speech and silence (Bortfeld et al. 2009, Dehaene-Lambertz et al. 2002, Pena et al. 2003). This early left lateralization has been confirmed using diffusion tensor imaging, a technique that is able to track white matter fascicles and myelination. The left hemisphere showed advanced development in 2-month-old infants (Dubois et al. 2008). Interestingly, those aspects of language processing that are usually right lateralized in adults, e.g., the processing of prosody, also appear to be right lateralized in infants (Homae et al. 2006, 2007).

In addition to this general lateralization pattern, recent results have allowed identification of the areas involved in language processing at a more fine-grained level. Gervain et al. (2008a), using near-infrared spectroscopy, have found that the newborn brain is able to extract identical, adjacent repetitions of syllables from speech stimuli. The repetitions were detected as some kind of perceptual Gestalt or primitive by the left (and to a lesser extent by the right) temporal areas immediately upon exposure. Over the course of the study, the repeated exposure to dozens of different stimuli, all instantiating the same underlying regularity (AAB: "mubaba," "penana," etc.), also gave rise to an increased response in the left frontal areas, suggesting the general pattern has been learned or extracted from the stimuli. This connection between the temporal areas, responsible for auditory processing, and the frontal areas, involved in higher-level learning and memory, has also been documented in a series of studies by Dehaene-Lambertz and her collaborators (Dehaene-Lambertz & Baillet 1998; Dehaene-Lambertz & Gliga 2004; Dehaene-Lambertz et al. 2006, 2008). These authors used activation speed to identify a circuit of areas, from the primary auditory cortex through the superior temporal gyrus to the inferior frontal area, which respond to speech in a hierarchical, cascading fashion, possibly integrating over increasingly large and/or abstract linguistic units.

These results indicate that brain organization shows structural and functional

specialization for language from the start. This is not to say, though, that language experience has no role to play. We demonstrated above how language experience shapes phonological and morphosyntactic development during the first year of life when measured behaviorally. In the past decade, numerous studies emerged documenting the underlying neural changes (for a recent review, see Kuhl & Rivera-Gaxiola 2008). For instance, Kuhl et al. (2008) found that at 7.5 months, better discrimination abilities for native phonemes, measured using electrophysiological techniques, correlate with the rate of later language development. This finding suggests that behavioral attunement to the native language is mediated by brain structures that become specifically responsive to frequently encountered, i.e., native, linguistic contrasts, which in turn promotes further learning of linguistic distinctions relevant for the native language and suppresses sensitivity to nonnative contrasts. Word learning also shows electrophysiological signatures at an early age. Familiar words evoke responses that are different in amplitude as well as in scalp distribution measurements from responses to unfamiliar words from about 9 months of age (Molfese 1990, Vihman et al. 2007).

Most of these studies were carried out with infants exposed to just one language. In many linguistic communities, though, exposure to multiple languages is the norm. An increasing body of research is now attempting to understand how such an environment affects phonological discrimination and categorization (Bosch & Sebastián-Gallés 1997, Conboy & Mills 2006, Mehler et al. 2004, Weikum et al. 2007, Werker & Byers-Heinlein 2008).

Interestingly, exposure to two languages from birth seems to affect development in other cognitive domains as well. In a series of experiments, Kovács & Mehler (2009a,b) have explored why bilingually raised children, having to learn twice as much about language as their monolingual peers, display a speed of acquisition comparable to that of monolingual infants. In the first study, Kovács & Mehler (2009a) compared 7-month-old monolingual and

bilingual groups in an eye-tracker task, where they had to learn to anticipate where a puppet would appear on the screen immediately after a trisyllabic word was heard. Both groups performed equally well in this task. During the second phase of the experiment, immediately after the first phase, both groups had to learn that the puppet appeared on the opposite side of the screen. Bilinguals learned this second task as fast as the first one, whereas monolinguals' performance was at chance. The authors concluded that continuous exposure to two languages during early infancy enhances the executive functions, attesting that the plasticity of certain brain regions prevents infants from potential confusion. In a second experiment with 12-month-olds, the same authors showed that when two structures, namely AAB and ABA, were used to cue infants to look to one side of the screen upon exposure to AAB and to the other side when ABA was heard (the presentation was interleaved), monolinguals learned to respond to the simpler structure AAB and were at chance for the other structure, whereas bilinguals learned both structures. The authors concluded, "The advantage of bilinguals may be related to the precocious development of control and selection abilities... This in turn may help them to learn more efficiently each of their languages. Such powerful learning abilities allow bilinguals to pass the linguistic milestones at the same rate as monolinguals" (Kovács & Mehler 2009b).

CONCLUSION

In this review, we presented theoretical approaches and underlying mechanisms proposed to explain infants' first steps into language. We have reviewed evidence suggesting that nativist and empiricist proposals are incomplete if they fail to include innate dispositions and learning in a broader, integrative, biologically anchored language acquisition theory. In addition, we have shown that a third type of mechanism, perceptual and memory constraints, needs to be evoked to provide a full account of early acquisition.

This integrative stance proposes that the three mechanisms are triggered by different properties of the input. For instance, statistical computations are evoked when the learner encounters an unsegmented speech stream. These computations selectively target some linguistic units, e.g., consonants, but not others, e.g., vowels. However, if the speech stream is already segmented, rule extraction and generalization mechanisms are used. In sum, the three processing and learning mechanisms complement as well as constrain each other.

Such an interaction of complementary mechanisms is not surprising from a biological point of view. Indeed, from an evolutionary perspective, the recruitment of a mechanism for a novel function is frequently observed (Jacob 1977). Therefore, it is plausible to assume that several of the mechanisms underlying our linguistic abilities are shared with other species. However, it remains true that only humans have language. Therefore, the quest is still on to identify the specific set of abilities that has emerged during our unique evolutionary history.

We have attempted to illustrate above how research into cognitive abilities and brain organization in young infants, in conjunction with information about the precursors that we share with other organisms, may shed light on the specifically human abilities that make us a language-learning animal.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Abercrombie D. 1967. *Elements of General Phonetics*. Edinburgh: Edinburgh Univ. Press

Abramson AS, Lisker L. 1970. Discriminability along the voicing continuum: cross-language tests. In *Proceedings of the Sixth International Congress of Phonetic Sciences*. Prague: Academia

Bahlmann J, Gunter TC, Friederici AD. 2006. Hierarchical and linear sequence processing: an electrophysiological exploration of two different grammar types. *J. Cogn. Neurosci.* 18(11):1829–42

Best CT, McRoberts GW. 2003. Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Lang. Speech* 46(2):183–216

Bonatti LL, Pena M, Nespor M, Mehler J. 2005. Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychol. Sci.* 16(6):451–59

Bortfeld H, Fava E, Boas DA. 2009. Identifying cortical lateralization of speech processing in infants using near-infrared spectroscopy. *Dev. Neuropsychol.* 34(1):52–65

Bosch L, Sebastián-Gallés N. 1997. Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition* 65(1):33–69

Bosch L, Sebastián-Gallés N. 2001. Early language differentiation in bilingual infants. In *Trends in Bilingual Acquisition*, ed. J Cenoz, F Genesee, pp. 71–93. Amsterdam: Benjamins

Brown RW. 1973. *A First Language: The Early Stages*. Cambridge, MA: Harvard Univ. Press

Chomsky N. 1957. *Syntactic Structures*. The Hague: Mouton

Chomsky N. 1959. A review of B.F. Skinner's verbal behavior. *Language* 35(1):26–58

Christophe A, Dupoux E, Bertoncini J, Mehler J. 1994. Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *J. Acoust. Soc. Am.* 95(3):1570–80

Church KW. 1987. Phonological parsing and lexical retrieval. *Cognition* 25:53–69

Conboy BT, Mills DL. 2006. Two languages, one developing brain: event-related potentials to words in bilingual toddlers. *Dev. Sci.* 9(1):F1–12

Csibra G, Gergely G. 2009. Natural pedagogy. *Trends Cogn. Sci.* 13(4):148–53

Curtin S, Fennell CT, Escudero P. 2009. Weighting of acoustic cues explains patterns of word-object associative learning. *Dev. Sci.* 12(5):725–31

Curtin S, Mintz TH, Byrd D. 2001. Coarticulatory cues enhance infants' recognition of syllable sequences in speech. *Proc. 25th Annu. Boston Univ. Conf. Lang. Dev.*, pp. 190–201. Somerville, MA: Cascadilla Press

Cutler A. 1994. Segmentation problems, rhythmic solutions. *Lingua* 92:81–104

Cutler A, Carter DM. 1987. The predominance of strong initial syllables in the English vocabulary. *Comput. Speech Lang.* 2:133–42

Dauer RM. 1983. Stress-timing and syllable-timing reanalyzed. *J. Phonetics* 11:51–62

Davis SM, Kelly MH. 1997. Knowledge of the English noun-verb stress difference by native and nonnative speakers. *J. Mem. Lang.* 36(3):445–60

Dehaene-Lambertz G, Baillet S. 1998. A phonological representation in the infant brain. *Neuroreport* 9(8):1885–88

Dehaene-Lambertz G, Dehaene S. 1994. Speed and cerebral correlates of syllable discrimination in infants. *Nature* 370(6487):292–95

Dehaene-Lambertz G, Dehaene S, Hertz-Pannier L. 2002. Functional neuroimaging of speech perception in infants. *Science* 298(5600):2013–15

Dehaene-Lambertz G, Gliga T. 2004. Common neural basis for phoneme processing in infants and adults. *J. Cogn. Neurosci.* 16:1375–87

Dehaene-Lambertz G, Hertz-Pannier L, Dubois J, Dehaene S. 2008. How does early brain organization promote language acquisition in humans? *Eur. Rev.* 16(4):399–411

Dehaene-Lambertz G, Hertz-Pannier L, Dubois J, Meriaux S, Roche A, et al. 2006. Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc. Natl. Acad. Sci. USA* 103(38):14240–45

Dubois J, Hertz-Pannier L, Cachia A, Mangin J, Le Bihan D, Dehaene-Lambertz G. 2008. Structural asymmetries in the infant language and sensori-motor networks. *Cereb. Cortex* 19:414–23

Eimas PD, Siqueland ER, Jusczyk PW, Vigorito J. 1971. Speech perception in infants. *Science* 171(968):303–6

Elman JL, Bates EA, Johnson MH, Karmiloff-Smith A, Parisi D, Plunkett K. 1996. *Rethinking Innateness: A Connectionist Perspective on Development.* Cambridge, MA: MIT Press

Endress AD, Dehaene-Lambertz G, Mehler J. 2007. Perceptual constraints and the learnability of simple grammars. *Cognition* 105(3):577–614

Endress AD, Mehler J. 2009. The surprising power of statistical learning: when fragment knowledge leads to false memories of unheard words. *J. Mem. Lang.* 60(3):351–67

Endress AD, Nespor M, Mehler J. 2009. Perceptual and memory constraints on language acquisition. *Trends Cogn. Sci.* In press

Endress AD, Scholl BJ, Mehler J. 2005. The role of salience in the extraction of algebraic rules. *J. Exp. Psychol.: Gen.* 134(3):406–19

Feher O, et al. 2009. De novo establishment of wild-type song culture in the zebra finch. *Nature* 459(28):564–68

Fenk-Oczlon G, Fenk A. 2005. Crosslinguistic correlations between size of syllables, number of cases, and adposition order. In *Sprache und Natürlichkeit, Gedenkband für Willi Mazerthaler*, ed. G Fenk-Oczlon, C Winkler, pp. 350–57. Tübingen: Narr

Fennell CT, Werker JF. 2003. Early word learners' ability to access phonetic detail in well-known words. *Lang. Speech* 46(2):245–64

Fennell CT, Werker JF. 2004. Infant attention to phonetic detail: knowledge and familiarity effects. *Proc. 28th Annu. Boston Univ. Conf. Lang. Dev.*, pp. 165–76. Boston, MA

Fiser J, Aslin RN. 2002a. Statistical learning of higher-order temporal structure from visual shape sequences. *J. Exp. Psychol.: Learn. Mem. Cogn.* 28(3):458–67

Fiser J, Aslin RN. 2002b. Statistical learning of new visual feature combinations by infants. *Proc. Natl. Acad. Sci. USA* 99(24):15822–26

Fitch WT, Hauser MD. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303(5656):377–80

Fitch WT, Hauser MD, Chomsky N. 2005. The evolution of the language faculty: clarifications and implications. *Cognition* 97(2):179–210; discussion 211–25

Friederici AD, Bahlomann J, Heim S, Schubotz RI, Anwander A. 2006. The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci. USA* 103(7):2458–63

Gallese V, Fadiga L, Fogassi L, Rizzolatti G. 1996. Action recognition in the premotor cortex. *Brain* 119:593–609

Gardner TJ, Naef F, Nottebohm F. 2005. Freedom and rules: the acquisition and reprogramming of a bird's learned song. *Science* 308(5724):1046–49

Gentner TQ, Hulse SH, Bentley GE, Ball GF. 2000. Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds. *J. Neurobiol.* 42(1):117–33

Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440(7088):1204–7

Gergely G, Csibra G. 2003. Teleological reasoning in infancy: the naive theory of rational action. *Trends Cogn. Sci.* 7(7):287–92

Gerken L, Landau B, Remez R. 1990. Function morphemes in young children's speech perception and production. *Dev. Psychol.* 26:204–16

Gerken L, McIntosh B. 1993. Interplay of function morphemes and prosody in early language. *Dev. Psychol.* 29(3):448–57

Gervain J, Macagno F, Cogoi S, Pena M, Mehler J. 2008a. The neonate brain detects speech structure. *Proc. Natl. Acad. Sci. USA* 105(37):14222–27

Gervain J, Nespor M, Mazuka R, Horie R, Mehler J. 2008b. Bootstrapping word order in prelexical infants: a Japanese-Italian cross-linguistic study. *Cogn. Psychol.* 57:56–74

Gleitman LR, Landau B. 1994. *The Acquisition of the Lexicon*. Cambridge, MA: MIT Press

Gopnik M, Crago MB. 1991. Familial aggregation of a developmental language disorder. *Cognition* 39:1–50

Grabe E, Low EL. 2002. Durational variability in speech and the rhythm class hypothesis. In *Papers in Laboratory Phonology*, ed. C Gussenhoven, N Warner, pp. 515–46. Berlin: Mouton de Gruyter

Guasti MT. 2002. *Language Acquisition: The Growth of Grammar*. Cambridge, MA: MIT Press

Haesler S, Rochefort C, Georgi B, Licznerski P, Osten P, Scharff C. 2007. Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus area X. *PLoS Biol.* 5(12):2885–97

Harris Z. 1955. From phoneme to morpheme. *Language* 31:190–222

Hauser MD, Chomsky NA, Fitch WT. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(5598):1569–79

Hauser MD, Newport EL, Aslin RN. 2001. Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition* 78(3):B53–64

Hay JS, Diehl RL. 2007. Perception of rhythmic grouping: testing the iambic/trochaic law. *Percept. Psychophys.* 69(1):113–22

Hayes B. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: Univ. Chicago Press

Hayes JR, Clark HH. 1970. Experiments in the segmentation of an artificial speech analogue. In *Cognition and the Development of Language*, ed. JR Hayes, pp. 221–34. New York: Wiley

Hochmann J-R, Azadpour M, Mehler. 2008. Do humans really learn AnBn artificial grammars from exemplars? *Cogn. Sci.* 32:1021–36

Höhle B, Weissenborn J. 2003. German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Dev. Sci.* 6(2):122–27

Homae F, Watanabe H, Nakano T, Asakawa K, Taga G. 2006. The right hemisphere of sleeping infant perceives sentential prosody. *Neurosci. Res.* 54(4):276–80

Homae F, Watanabe H, Nakano T, Taga G. 2007. Prosodic processing in the developing brain. *Neurosci. Res.* 59(1):29–39

Iversen JR, Patel AD, Ohgushi K. 2008. Perception of rhythmic grouping depends on auditory experience. *J. Acoust. Soc. Am.* 124(4):2263–71

Jacob F. 1977. Evolution and tinkering. *Science* 196(4295):1161–66

James AL. 1940. *Speech Signals in Telephony*. London: Sir Isaac Pitman & Sons

Johnson EK, Jusczyk PW. 2001. Word segmentation by 8-month-olds: when speech cues count more than statistics. *J. Mem. Lang.* 44(4):548–67

Jusczyk PW, Hohne EA, Bauman A. 1999. Infants' sensitivity to allophonic cues to word segmentation. *Percept. Psychophys.* 61:1465–76

Jusczyk PW, Houston DM, Newsome MR. 1999. The beginnings of word segmentation in English-learning infants. *Cogn. Psychol.* 39(3):159–207

Knops A, Thirion B, Hubbard EM, Michel V, Dehaene S. 2009. Recruitment of an area involved in eye movements during mental arithmetic. *Science* 324(5934):1583–85

Kovács AM, Mehler J. 2009a. Cognitive gains in 7-month-old bilingual infants. *Proc. Natl. Acad. Sci. USA* 106(16):6556–60

Kovács AM, Mehler J. 2009b. Flexible learning of multiple speech structures in bilingual infants. *Science.* In press

Kučera H, Francis WN. 1967. *Computational Analysis of Present-Day American English*. Providence, RI: Brown Univ. Press

Kudo N, Nonaka Y, Mizuno K, Okanoya K. 2006. *Statistical learning and word segmentation in neonates: an ERP evidence*. Presented at Annu. Meet. XVth Bienn. Int. Conf. Infant Studies, Kyoto, Japan

Kuhl PK. 2004. Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5(11):831–43

Kuhl PK, Conboy BT, Coffey-Corina S, Padden D, Rivera-Gaxiola M, Nelson T. 2008. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363(1493):979–1000

Kuhl PK, Rivera-Gaxiola M. 2008. Neural substrates of language acquisition. *Annu. Rev. Neurosci.* 31:511–34

Kuhl PK, Stevens E, Hayashi A, Deguchi T, Kiritani S, Iverson P. 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Dev. Sci.* 9(2):F13–21

Ladefoged P. 1993. *A Course in Phonetics.* New York: Harcourt Brace Jovanovich

Lenneberg EH. 1967. *The Biological Foundations of Language*. New York: Wiley

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. 1967. Perception of the speech code. *Psychol. Rev.* 74(6):431–61

Liberman AM, Harris KS, Hoffman HS, Griffith BC. 1957. The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54:358–68

Marcus GF, Vijayan S, Rao SB, Vishton PM. 1999. Rule learning by seven-month-old infants. *Science* 283(5398):77–80

Marler P, Peters S. 1981. Sparrows learn adult song and more from memory. *Science* 213:780–82

Mattys SL, Jusczyk PW. 2001a. Do infants segment words or recurring contiguous patterns? *J. Exp. Psychol.: Hum. Percept. Perform.* 27(3):644–55

Mattys SL, Jusczyk PW. 2001b. Phonotactic cues for segmentation of fluent speech by infants. *Cognition* 78(2):91–121

Mattys SL, Jusczyk PW, Luce PA, Morgan JL. 1999. Phonotactic and prosodic effects on word segmentation in infants. *Cogn. Psychol.* 38(4):465–94

Maye J, Werker JF, Gerken L. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82(3):B101–111

Mehler J. 1974. Apprendre par désapprendre. In *L'Unité de l'homme: essais et discussions*, ed. E Morin, M Piatelli-Palmarini, pp. 187–319. Paris: Éditions du Seuil

Mehler J, Bertoncini J, Barriere M, Gerschenfeld DJ. 1978. Infant recognition of mother's voice. *Perception* 7:491–97

Mehler J, Dupoux E. 1994. *What Infants Know: The New Cognitive Science of Early Development*. Cambridge, MA: Blackwell

Mehler J, Jusczyk PW, Lambertz G, Halsted N, Bertoncini J, Amiel-Tison C. 1988. A precursor of language acquisition in young infants. *Cognition* 29:143–78

Mehler J, Sebástian-Gallés N, Nespor M. 2004. Biological foundations of language: language acquisition, cues for parameter setting and the bilingual infant. In *The New Cognitive Neuroscience*, ed. M Gazzaniga, pp. 825–36. Cambridge, MA: MIT Press. 3rd ed.

Miller JE, et al. 2008. Birdsong decreases protein levels of FoxP2, a molecule required for human speech. *J. Neurophysiol.* 100(4):2015–25

Mintz TH. 2002. Category induction from distributional cues in an artificial language. *Mem. Cogn.* 30(5):678–86

Molfese DL. 1990. Auditory evoked responses recorded from 16-month-old human infants to words they did and did not know. *Brain Lang.* 38(3):345–63

Moon C, Cooper RP, Fifer WP. 1993. Two-day-olds prefer their native language. *Infant Behav. Dev.* 16(4):495–500

Morgan JL, Demuth K. 1996. *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition.* Hillsdale, NJ: Erlbaum

Morgan JL, Shi R, Allopenna P. 1996. Perceptual bases of rudimentary grammatical categories: toward a broader conceptualization of bootstrapping. In *Signal to Syntax*, ed. JL Morgan, K Demuth, pp. 263–83. Hillsdale, NJ: Erlbaum

Morse PA, Molfese D, Laughlin NK, Linnville S, Wetzel F. 1987. Categorical perception for voicing contrasts in normal and lead-treated rhesus monkeys: electrophysiological indices. *Brain Lang.* 30(1):63–80

Nazzi T, Bertoncini J. 2003. Before and after the vocabulary spurt: two modes of word acquisition? *Dev. Sci.* 6(2):136–42

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(3):756–66

Nespor M. 1990. On the rhythm parameter in phonology. In *Logical Issues in Language Acquisition*, ed. I Roca, pp. 157–75. Dordrecht: Foris

Nespor M, Guasti MT, Christophe A. 1996. Selecting word order: the rhythmic activation principle. In *Interfaces in Phonology*, ed. U Kleinhenz, pp. 1–26. Berlin: Akademie Verlag

Nespor M, Pena M, Mehler J. 2003. On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio* 2:203–31

Nespor M, Shukla M, van de Vijver R, Avesani C, Schraudolf H, Donati C. 2008. Different phrasal prominence realization in VO and OV languages. *Lingue e Linguaggio* 7(2):1–28

Newport EL, Aslin RN. 2004. Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cogn. Psychol.* 48(2):127–62

Newport EL, Hauser MD, Spaepen G, Aslin RN. 2004. Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cogn. Psychol.* 49(2):85–117

Onishi KH, Baillargeon R. 2005. Do 15-month-old infants understand false beliefs? *Science* 308(5719):255–58

Pena M, Bonatti LL, Nespor M, Mehler J. 2002. Signal-driven computations in speech processing. *Science* 298(5593):604–7

Pena M, Maki A, Kovacic D, Dehaene-Lambertz G, Koizumi H, et al. 2003. Sounds and silence: an optical topography study of language recognition at birth. *Proc. Natl. Acad. Sci. USA* 100(20):11702–5

Perruchet P, Rey A. 2005. Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychon. Bull. Rev.* 12(2):30713

Pike KL. 1945. *The Intonation of American English*. Ann Arbor: Univ. Mich. Press

Pinker S. 1984. *Language Learnability and Language Development*. Cambridge, MA: Harvard Univ. Press

Pinker S, Jackendoff R. 2005. The faculty of language: What's special about it? *Cognition* 95(2):201–36

Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R. 2009. Neural correlates of categorical perception in learned vocal communication. *Nat. Neurosci.* 12(2):221–28

Prather JF, Peters S, Nowicki S, Mooney R. 2008. Precise auditory-vocal mirroring in neurons for learned vocal communication. *Nature* 451:305–10

Quine WV. 1960. *Word and Object*. Cambridge, MA: MIT Press

Ramon-Casas M, Swingley D, Sebástian-Gallés N, Bosch L. 2009. Vowel categorization during word recognition in bilingual toddlers. *Cogn. Psychol.* 59(1):96–121

Ramus F. 2002. Acoustic correlates of linguistic rhythm: perspectives. In *Proceedings of Speech Prosody 2002*, ed. B Bel, I Marlien, pp. 115–20. Aix-en-Provence, France

Ramus F, Hauser MD, Miller C, Morris D, Mehler J. 2000. Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science* 288(5464):349–51

Ramus F, Mehler J. 1999. Language identification with suprasegmental cues: study based on speech resynthesis. *J. Acoust. Soc. Am.* 105(1):512–21

Ramus F, Nespor M, Mehler J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3):265–92

Redington M, Chater N, Finch S. 1998. Distributional information: a powerful cue for acquiring syntactic categories. *Cogn. Sci.* 22(4):425–69

Rizzolatti G, Fogassi L, Gallese V. 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2(9):661–70

Saffran JR, Aslin RN, Newport EL. 1996. Statistical learning by 8-month-old infants. *Science* 274(5294):1926–28

Saffran JR, Hauser M, Seibel R, Kapfhamer J, Tsao F, Cushman F. 2008. Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition* 107(2):479–500

Saffran JR, Thiessen ED. 2003. Pattern induction by infant language learners. *Dev. Psychol.* 39(3):484–94

Sansavini A, Bertoncini J, Giovanelli G. 1997. Newborns discriminate the rhythm of multisyllabic stressed words. *Dev. Psychol.* 33(1):3–11

Scott LS, Pascalis O, Nelson CA. 2007. A domain-general theory of the development of perceptual discrimination. *Curr. Dir. Psychol. Sci.* 16(4):197–201

Senghas A, Kita S, Özyürek A. 2004. Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science* 305(5691):1779–82

Shannon CE. 1948. A mathematical theory of communication. *Bell Syst. Tech. J.* 27:379–423, 623–56

Shi R, Cutler A, Werker J, Cruickshank M. 2006. Frequency and form as determinants of functor sensitivity in English-acquiring infants. *J. Acoust. Soc. Am.* 119(6):EL61–67

Shi R, Werker JF. 2001. Six-month-old infants' preference for lexical words. *Psychol. Sci.* 12(1):71–76

Shi R, Werker JF, Morgan JL. 1999. Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition* 72(2):B11–21

Shipley EF, Smith CS, Gleitman LR. 1969. A study in the acquisition of language: free responses to commands. *Language* 45:322–42

Shukla M, Nespor M, Mehler J. 2007. An interaction between prosody and statistics in the segmentation of fluent speech. *Cogn. Psychol.* 54(1):1–32

Slobin DI. 1997. *The Crosslinguistic Study of Language Acquisition.* Hillsdale, NJ: Erlbaum

Stager CL, Werker JF. 1997. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388(6640):381–82

Swingley D. 2005. Statistical clustering and the contents of the infant vocabulary. *Cogn. Psychol.* 50(1):86–132

Swingley D. 2009. Contributions of infant word learning to language development. *Philos. Trans. Royal Soc. B.* In press

Taga G, Asakawa K. 2007. Selectivity and localization of cortical response to auditory and visual stimulation in awake infants aged 2 to 4 months. *NeuroImage* 36(4):1246–52

Teinonen T, Fellman V, Naatanen R, Alku P, Huotilainen M. 2009. Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* 10:21

Thiessen ED. 2007. The effect of distributional information on children's use of phonemic contrasts. *J. Mem. Lang.* 56(1):16–34

Tomasello M. 2000. Do young children have adult syntactic competence? *Cognition* 74(3):209–53

Toro JM, Nespor M, Mehler J, Bonatti LL. 2008. Finding words and rules in a speech stream: functional differences between vowels and consonants. *Psychol. Sci* 19(2):137–44

Toro JM, Trobalon JB. 2005. Statistical computations over a speech stream in a rodent. *Percept. Psychophys.* 67(5):867–75

Vihman MM, Thierry G, Lum J, Keren-Portnoy T, Martin P. 2007. Onset of word form recognition in English, Welsh, and English-Welsh bilingual infants. *Appl. Psycholinguistics* 28(3):475–93

Vouloumanos A, Werker JF. 2004. Tuned to the signal: the privileged status of speech for young infants. *Dev. Sci.* 7(3):270

Vouloumanos A, Werker JF. 2007. Listening to language at birth: evidence for a bias for speech in neonates. *Dev. Sci.* 10(2):159–64

Waxman SR, Gelman SA. 2009. Early word-learning entails reference, not merely associations. *Trends Cogn. Sci.* 13(6):258–63

Weikum WM, Vouloumanos A, Navarra J, Soto-Faraco S, Sebastián-Gallés N, Werker JF. 2007. Visual language discrimination in infancy. *Science* 316(5828):1159

Weissenborn J, Höhle B, Kiefer D, Cavar D. 1996. Children's sensitivity to word-order violations in german: evidence for very early parameter-setting. *Proc. 22nd Annu. Boston Univ. Conf. Lang. Dev*, pp. 756–77. Boston, MA

Werker JF, Byers-Heinlein K. 2008. Bilingualism in infancy: first steps in perception and comprehension. *Trends Cogn. Sci.* 12(4):144–51

Werker JF, Tees RC. 1984a. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7(1):49–63

Werker JF, Tees RC. 1984b. Phonemic and phonetic factors in adult cross-language speech perception. *J. Acoust. Soc. Am.* 75(6):1866–78

Yang CD. 2004. Universal grammar, statistics or both? *Trends Cogn. Sci.* 8(10):451–56

Yang C, Gambell T. 2004. Statistics learning and universal grammar: modeling word segmentation. In *COLING 2004: Psycho-Computational Models of Human Language Acquisition*, pp. 51–54. Geneva, Switzerland

Yip MJ. 2006. The search for phonology in other species. *Trends Cogn. Sci.* 10(10):442–46

Yoshida KA, Fennell CT, Swingley D, Werker JF. 2009. Fourteen-month-old infants learn similar sounding words. *Dev. Sci.* 12(3):412–18

Yoshida KA, Iversen JR, Patel AD, Mazuka R, Nito H,et al. 2009. The development of perceptual grouping biases in infancy. *Cognition*. Manuscript under revision

**Annual Review of
Psychology**

Volume 61, 2010

Contents

Errata

An online log of corrections to *Annual Review of Psychology* articles may be found at
http://psych.annualreviews.org/errata.shtml