8

Why is language unique to humans?

JACQUES MEHLER, MARINA NESPOR, AND MARCELA PEÑA

8.1 Introduction

Linguists, psychologists, and neuroscientists have studied language acquisition with the tools and models available to their respective fields. Linguists elaborated some of the most sophisticated theories to account for how this unique human competence arises in the infants' brains. Chomsky (1980) formulated the parameter setting theory (hereafter, PS) to account for how infants, on the basis of partial and noisy language input, acquire grammar. PS assumes that infants are born with "knowledge" of Universal Grammar (UG). This includes both genetically determined universal principles and binary parameters. Universal principles describe the properties common to all natural languages. Binary parameters capture the grammatical properties on which natural languages differ from one another. The linguistic input determines the particular value of a parameter. PS postulates that exposure to the surrounding language determines how the parameters of UG are set.[1]

We acknowledge that PS has many virtues. It addresses the problem of language acquisition without making unjustified but common simplifications, for example, that imitation is the privileged mechanism responsible for the emergence of linguistic competence. The theory, furthermore, is quite appealing because it assumes, realistically, a biological perspective, namely, that the child is equipped with a species-specific mechanism to acquire natural language. Moreover, the PS theory has been formulated with sufficient detail and precision as to make it easy to falsify. In contrast, proposals that assume that language is acquired by means of a general learning device appear more difficult to support. Criticisms of proposals according to which general learning mechanisms are sufficient to explain language acquisition have been given by many

Topics in Integrative Neuroscience: From Cells to Cognition, ed. James R. Pomerantz. Published by Cambridge University Press.

theoreticians (see Chomsky, 1959; Fodor, 1975; Lenneberg *et al.*, 1964; Pinker, 1984).

All theories agree that at least parts of grammar have to be learned. What distinguishes the different positions is the scope and nature of learning. How does the learning proceed? PS assumes an initial state characterized by knowledge specific to language. In contrast, theoreticians who favor a general learning mechanism, assume that the initial state is characterized by learning principles that apply to all areas in which the organism gains knowledge. PS has the advantage that it is rather easy to falsify. If syntax cannot be acquired given the normal input, then PS would have to be abandoned. Indeed, if PS turns out to be misguided, badly informed, or incorrect, another theory will have to be formulated and evaluated. This is far from being an exceptional situation. Rather, it is one that is obtained in all scientific domains. In contrast, recent generic learning accounts (see Plunkett, McClelland, and many others) have not yet been presented in sufficient detail to be falsifiable. In this chapter, we ignore the generic learning account and focus on some aspects of PS.

So far, we have highlighted the positive aspects of PS. However, a problem resides in the hidden assumptions that investigators have made when trying to explicate the learning of grammar. PS was formulated with syntax acquisition in mind and investigators generally assumed that infants have already gained, in one way or another, knowledge of the lexicon, including the phonological information it carries, before setting grammatical parameters. If this were the case, both the lexical and the phonological properties of the language could be learned without having to consider syntax. Only if one believes that infants store the sounds in the surrounds ignoring any additional information that they contain, could one understand why researchers interested in the acquisition of syntax have ignored the first year of life: during this period, babies would only memorize the sounds in their surrounds and acquire the first few words. If one makes such presuppositions, it seems also reasonable to assume that infants set grammatical parameters after acquiring a basic vocabulary. By and large, scholars working in the PS tradition have assumed that the first year of life can be neglected without missing essential aspects of the acquisition process. In fact, if syntax is the crux of language and there is nothing syntactic being learned during the first year of life, why should one study that period at all? The presupposition that acquisition starts with the first linguistic productions, roughly at 8 months of age or later, explains why PS investigators have exclusively reported data on language production.

Supporting the PS position may have been justified by data on animal behavior. Indeed, animals with auditory systems similar to our own tend to respond to speech patterns much like infants younger than 8 months (see Kuhl, 1987;

Ramus *et al.*, 2000 among many others). Apes, but also dogs, have "lexicons" that can attain a few dozen words (see Premack, 1971, 1986). Tamarins, chinchillas, and several other animals treat and respond to sounds much like humans (see Doupe, this book). However, their perceptual and mnemonic abilities are not sufficient to enable them to construct a grammar comparable to that of human languages. In contrast to the assumption that the first years of life is irrelevant to the acquisition of syntax, we show below that language acquisition begins with the onset of life. Indeed, recent data supports the view that the sound pattern of language plays an important role in the learning of syntax.

Psychologists have explored general learning accounts of knowledge acquisition, including language. Most of those studies have tried to understand how productive a model of language acquisition entirely based on associations can be. Within this stream of research, the brain is regarded as a huge network that works in a Hebbian fashion.[2] This explains why many psychologists, as well as many neuroscientists, though by no means all, have adopted a contrasting viewpoint from that of linguists. Their tendency has been to neglect syntax and assume that by focusing exclusively on speech perception and production, a functional theory of language will ensue. Undeniably, behavioral scientists have achieved great success studying perception and production. Some of them believe that it is sufficient to study how language production and perception unfold during development to understand how syntax (or semantics) is computed by the mind. This stance was strengthened because, while it is easy to study how babies or animals perceive speech sounds, it is very hard to study the acquisition of syntax in the laboratory. Psychologists who work assuming a generic learning mechanism behave as if the mystery of syntax acquisition will disappear by observing how infants learn to conform to the structure of language (see Seidenberg & MacDonald, 1999; Tomasello, 2000, among many others).

We believe that true progress will be accomplished once the above divide of research strategies is overcome. Losing sight of the uniqueness of syntax is dangerous and so is neglecting how signals are processed and represented by the very young infant. Indeed, the linguistic input can be viewed as speech signal (or hand gestures for the deaf) that contains information about different aspects of grammar, syntax included; that is, the triggers of different parameters may be present in some shallow acoustic form in the input the child receives. However, unlike many reflexes that are triggered by a sensorial stimulus even the first time the organism encounters it, it seems highly probable that speech signals do not trigger the setting of a syntactic parameter the first time the infant listens to a sentence. Rather, it seems more likely that the child would gather enough information to draw a conclusion about the appropriate value of a parameter. In fact, many infants (maybe even a majority) are exposed to two languages from birth onwards. The two languages might require that a single parameter be set in two different ways. Will the information that is necessary to fix a parameter in one language of exposure be masked by noise from the other language? Will there be two files, one for each of the two languages? Or rather will there be a single noisy file that will result in utter confusion to the child? These are some of the issues that are essential for linguists and cognitive neuroscientists to confront together to bring their theoretical stances in closer harmony with one another and with the facts.

Fortunately, the polarity we described above is already diminishing. The interaction between the fields began to increase when scholars began to realize that grammar acquisition, even in a tradition like the one defined by PS, remains rather vague. Indeed, even though linguists studied the influence of syntax on the phonological shape of speech (see Nespor & Vogel, 1986; Selkirk, 1984, among others), they have not explored how speech signals trigger the fixation of parameters in infants. As will be argued below, we believe that the time is ripe to explore how humans sample information from the surrounds to discover the abstract properties of language. Only then will we be able to understand what the essential difference is between the human and the evolved ape's brain. That will be the time when a new impulse will be given to the study of the biological foundations of language.

Studies by Chomsky (1980, 1986), Wexler and Culicover (1980), Pinker (1984), and others have lucidly argued for a PS conception of language acquisition. However, the PS formulation may have been seriously under-specified making it hard to judge its adequacy. In fact, Mazuka (1996) has argued that, in its usual formulation, PS contains a fatal paradox. Of course, solutions to most of these problems might turn up in the years to come. Morgan *et al.* (1987), Cutler (1994), and Nespor *et al.* (1996), among others, have proposed some putative solutions. However, few proposals have explored how the infant evaluates and computes the triggering signals. Some recent results suggest that nearly 2-month-olds are sensitive to the prosodic correlates of the different values of the head-complement parameter (Christophe *et al.*, 1997; Christophe *et al.*, 2003).

In the early 1980s, some psychologists and some linguists like Wanner and Gleitman (1982) already foresaw some of the difficulties in existing theories of grammar acquisition and proposed that phonological bootstrapping may help the infant out of its quandary. Wanner and Gleitman (1982) held that some properties of the phonological system that the child is learning may help uncover lexical and syntactic properties. Some years later, Morgan and Demuth (1996) added that specifically prosody might contain signals that can act as triggers helping the child to learn syntax. Indeed, these authors conclude, as we do above, that the study of the speech signals that can act as triggers is essential to understand the first steps into language. A better understanding of the speech signal might also uncover whether PS is a solution to the problem highlighted by learnability theorists: the poverty of the stimulus (see Wexler & Culicover, 1980; and many others). The postulation of innate structure was the way chosen to overcome the poverty of the stimulus problem. Today, we see that this proposal is not sufficiently specific. Indeed, if an important part of the endowment comes as binary parameters, we still need to understand how these are set to the values adequate for the surrounding language. The general assumption was that by understanding a few words, simple sentences like *drink the juice*, *eat the soup*, will allow the child to generalize the fact that, in his/her language, objects follow verbs. As Mazuka (1996) pointed out, this assumption is unwarranted. Indeed, how does the child know that soup means *soup* (Noun) rather than *eat* (Verb)? Even if Mom always says *eat* in front of diverse foods, the child could understand that what she means is *food*! If the signals were to inform the child about word order, one could find a way out of this paradox. Before we know if this is a true solution, we need to ask whether such signals exist and if they do, whether the infant can process them.

The prosodic bootstrapping hypothesis arose from linguistic research that focused on the prosodic properties that are systematically associated with specific syntactic properties (see Nespor & Vogel, 1986; Selkirk, 1984, among many others). These authors found interesting associations between these two grammatical levels, making plausible the notion that signals might cue the learner to postulate syntactic properties in an automatic, encapsulated fashion.

Let us assume that babies are born with Universal Grammar. It still is essential to understand how they learn their maternal language. We know that the properties of the speech signals are processed very precociously; and if one believes, as we do, that speech signals contain the information that is necessary to set the main parameters, we still have to explain what happens during the first 18 months of life. What is the baby doing that takes it so long to get going? What is the infant learning throughout this period? Since infants perceive the cues that can set triggers and since these are supposed to function in an automatic and encapsulated way, we are committed to the view that infants have "learned" many aspects of the language before they begin to produce speech. We have the responsibility, however, to give an account of the specific processes that happen during the first months of life. As we argued above (p. 4), a parameter will not be set after listening to a single utterance. Rather, properties of utterances are stored and only when the information becomes "reliable" will it be used to set a parameter. Since some parameters can only be set after other grammatical properties have already been acquired (and each of them requiring considerable information storage), we might understand the "slow" pace of learning. Learning the outstanding properties of grammar is just one aspect of language acquisition. In addition, the child has to learn a great deal of arbitrary linguistic properties. The sound of words is arbitrary. One should also not forget that most words are heard in connected speech. Thus, we must investigate how the infant parses the input to identify words. A proposal made by Saffran *et al.* (1996) is that this requires the inspection of the statistical properties of the incoming speech signals.

Let us now spell out the purpose of the present chapter. While we assume that UG is part of the infant's endowment and that it guides language acquisition, we also acknowledge that statistical properties of the language spoken in the surrounds inform and guide learning. This is in contrast to the position of some theorists as MacWhinney (1987) and Seidenberg and MacDonald (1999) who argue that it is unnecessary to pay attention to grammar learning, since all that is required is to explain how the child learns to comprehend and produce language. The authors, and many others, believe that it is possible to explain linguistic performance exclusively on the basis of the infant's sensitivity to the statistical properties of signals. Generally, this position is defended on the basis of rather simplified scenarios in which each solution is proposed for the acquisition of just one aspect of grammar. How would their model stand up in the real setting in which infants learn language, not to mention bilingual settings, or the creolization of pidgin languages.

The above presentation makes it clear that more data and research are needed to understand how the biological human endowment interacts with the learning abilities during the first months of life. We are in a rather good position, because during the last few years, new and fascinating results have been secured allowing us to start having a coherent picture of language acquisition.

8.2 Innate dispositions for language?

Before and after birth, infants experience speech in noisy environments. A conjecture that is often made by pediatricians and naïve observers is that this cacophony that infants experience is not a problem because they had learned to attend to speech during gestation. The womb, however, is not such a quiet place. Indeed, experiments carried out with pregnant quadrupeds and also on volunteer pregnant women reveal that intra-uterine noise tends to be even more important than the noise that the infant encounters after birth. The bowels, blood circulation, other body movements, to mention a few sources, generate noise with considerable energy (Querleu *et al.*, 1988). Thus, acoustic stimulation in the womb will not explain how infants segregate speech from background noise. How does the infant identify the signals that carry linguistic information? Why are music, telephone rings, animal sounds, etc. segregated during language acquisition?

Psycholinguists have explored experimentally this difficult question. Colombo and Bundy (1983) have reported that infants respond preferentially to speech streams as compared to other noises. This result, however, is difficult to evaluate, since it is always conceivable that infants would prefer a nonspeech stimulus different from the one used by Colombo and Bundy (1983). Maybe, a melody might be found that is equally attractive as the speech stream. Few experimenters have explored this question in a more convincing way. Mehler *et al.*, 1988 found that neonates behave differently when they are exposed to normal utterances as compared to the same utterances played backward. These authors interpret their finding as showing that infants attend to speech rather than to other stimuli even when they are matched for pitch, intensity, and duration.

More evidence is needed to be convinced that the neonate's brain responds specifically to speech sounds rather than to the human voice (regardless of whether it is producing speech or coughs, cries, sneezes, etc.). Humans are incapable to produce backward speech. The impossibility of the vocal tract to produce backward speech might be an alternative explanation of Mehler *et al.*'s results mentioned above. The contrast between a natural utterance (producible by the human vocal tract) and a machine-made rearrangement of the same utterance (that no human vocal tract could produce) may be the relevant factor, rather than the contrast between speech and nonspeech that the authors invoke. Belin *et al.* (2000) have recently claimed that there is a brain area that is devoted to processing conspecific vocal productions. He examined adult subjects in an fMRI imaging experiment while they were listening to various speech and nonspeech sounds all made by the human vocal tract (i.e., speech but also laughs, sighs, and various onomatopoeia). In response to all these stimuli, he found bilateral activation along the upper banks of the STS (Superior Temporal Sulcus). However, vocal sounds elicit greater activation than nonvocal sounds bilaterally in nonprimary auditory cortex. If Belin *et al.*'s results are corroborated, one might explain the speech vs. backward speech results mentioned above because only speech can be produced by the human vocal tract.

Belin and his colleagues have argued that the brain is organized to process human voices much like other parts of the brain are organized to process human faces. Indeed, Kanwisher *et al.* (1997) have proposed that faces are processed in a specific area, the FFA (fusiform face area). According to Belin *et al.*, human voice is processed in the STS. This conclusion may be premature since we do not yet know the set of stimuli that activate the voice recognition area.[3]

Our own outlook is that it is essential to study the specificity of cortical areas devoted to process different information types, before any prior learning has occurred. Establishing whether certain areas of the brain are organized in specific ways is essential for the study of infancy and also for the construction of theories of development. Thus, in contrast to the above-described investigations our research focuses mainly on the initial state of the cognitive system. Adults may have already learned how to process and encode faces or human vocal tract production, and as a result have taken possession of cortical tissue for this purpose. Therefore, to distinguish what is due to our endowment and what arises as a consequence of learning, it is necessary to investigate very young infants and whenever possibly neonates since in the first months many acquisitions have already been documented (for some investigations that bear mostly on language, see Jusczyk, 1997; Kuhl *et al.*, 1992; Mehler & Dupoux, 1994; Werker & Tees, 1984).

Standard neurological science has gathered evidence that the left hemisphere (LH) is more involved with language representation and processing than the right hemisphere (RH). Are infants born with specific LH areas devoted to speech processing or is the LH specialization the sole result of experience? The response to this question is still tentative. Numerous investigations have reported that infants are born with speech processing abilities similar to those displayed by experienced adults. For instance, infants discriminate all the phonetic contrasts that arise in natural languages, (see Jusczyk, 1997; Mehler & Dupoux, 1994). At first, this finding was construed as showing that humans are born with specific neural machinery devoted to speech. Subsequent investigations, however, demonstrated that basic acoustic processes are sufficient to explain these early abilities that humans share with other organisms (see Jusczyk, 1997; Jusczyk *et al.*, 1977; Kuhl & Miller, 1975). Thus, it is reasonable to postulate a species-specific disposition to acquire natural language, but we still lack data to ground the view that we are born with cortical structures specifically dedicated to the processing of speech.

As we mentioned above, functional asymmetries, in particular a superiority of the LH, seem to be related to speech processing. A great deal of neuropsychological evidence points in that direction (see Bryden & Allard, 1981; Dronkers, 1996; Geschwind, 1970). Likewise, experimental studies carried out on normal adult volunteers suggest that LH dominance characterizes speech processing (see Bertelson, 1982 among many others). We still ignore whether such LH superiority is the consequence of language acquisition or whether language is mastered because of this tissue specialization. Developmental psychologists investigated this issue in some detail. Most behavioral studies found an asymmetry in very young humans (see Bertoncini *et al.*, 1989; Best *et al.*, 1982; Segalowitz & Chapman, 1980). A few ERP studies have also found trends for LH superiority in young infants (see Dehaene-Lambertz & Dehaene, 1994; Molfese & Molfese, 1979). Both the behavioral and the ERP data suggest that LH superiority exists in the infant's brain but more evidence is desirable to strengthen and to further understand the cortical organization of the immature brain. Fortunately, we are entering a new era and it is becoming possible to use more advanced imaging methods to study the functional brain organization in newborn infants. A number of methods are being pursued in parallel. Numerous groups have begun to study healthy infants using fMRI (G. Dehaene-Lambertz, personal communication). In the following section, we focus on recent results we obtained with Optical Topography (OT).

8.3 Brain specialization in newborns: evidence from OT

Optical Topography is a method derived from Near Infrared technology developed in the early 1950s (see Villringer & Chance, 1997 for an excellent review of the field). This technology allows us to estimate the vascular response of the brain following stimulation.[4] In particular, it allows to estimate the concentration of oxyhemoglobin (oxyHb) and deoxyhemoglobin (deoxyHB) over a given area of the brain.

We used a prototype device produced by Hitachi and modified by us. This device allowed us to place two sets of optic fibers on each side of the infant's head. We first studied the simultaneous activation of two areas of the brain. These areas were located to be, as nearly as possible, homologous to each other on the LH and the RH. We assume that we have placed the probes so as to measure activity over the RH and the LH temporal and parietal areas. Each infant was tested with three kinds of blocks of stimuli. In one condition (Forward Speech, FW), infants hear sequences of 15 seconds of connected French utterances separated from one another by periods of silence of variable duration (from 25 to 35 seconds). In another condition (Backward Speech, BW), infants are tested like in the FW condition but with the speech sequences played backward (the signal was converted from FW to BW using a speech editor). Ten such blocks are presented in the FW and in the BW conditions for each infant. Finally, in another condition, infants are exposed to silence for a duration comparable to the average duration of the above conditions. The latter is a comparison measure for the other two conditions.

Not all infants completed the ten blocks in each condition. In order for an infant to be kept in the final data analysis, the subject had to complete at least

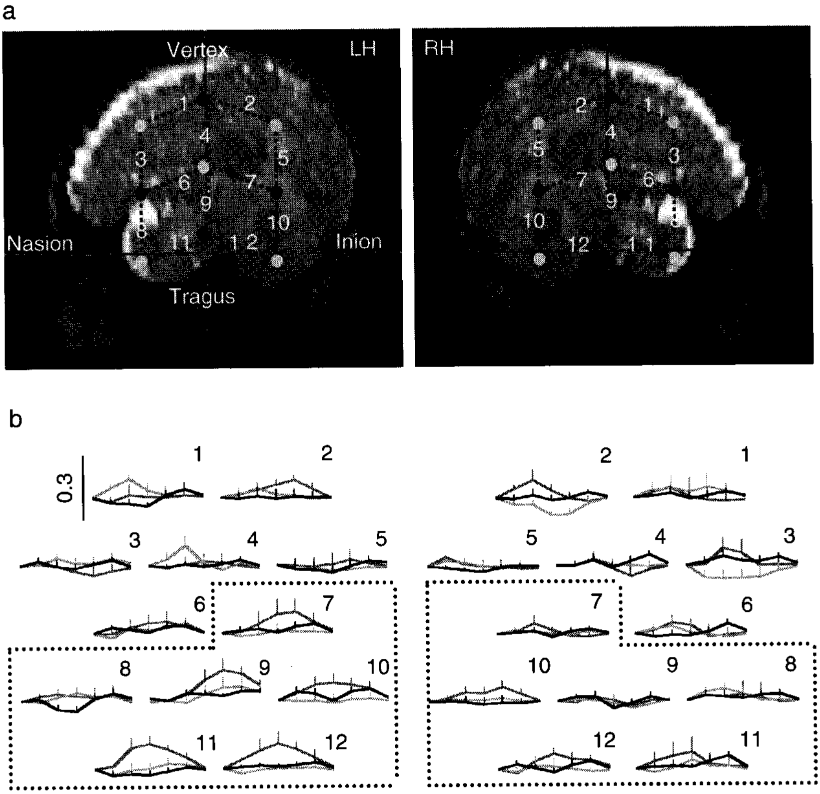

Figure 8.1 Positioning of the OT probes and observed results. (a) OT channels projected on an MR image of a 2-month-old Infant. Red dots correspond to emitter and blue dots to detector optical fibers. The numbers on the black dotted lines, between adjacent emitter–detector pairs of fibers, correspond to the channels from which changes in Hb concentration were estimated. Indicated skull landmarks (inion, nasion, tragus, and vertex) were used to place the probes. (b) The numbers above the plots correspond to channel numbers in a. The plots show the grand average of the mean of total Hb (mmol.mm) for successive 5-s windows. The first window begins 5 s before the onset of a block. The vertical black line in channel 1 of the LH indicates the range of total Hb concentration in mmol.mm valid for all of the channels. Total Hb is plotted in red for FW, in light green for BW, and in blue for SIL. Ascending bars indicate SDs. The six channels enclosed within dotted lines (7–12) cover the temporal regions below the Sylvian fissure (lower channels). Channels 1–6 were placed over the frontoparietal regions above the Sylvian fissure (upper channels); (with permission from PNAS) (for color image please see plate section).

three blocks in each one of the three conditions – FW, BW, and Silence. The preliminary results suggest that like in adults, the hemodynamic response begins 4 to 5 seconds after the infant receives the auditory stimulation. This time-locked response appears more clearly for the oxyHB than for the deoxyHB. The pattern of results shows that roughly 5 seconds after the presentation of the

FW utterances, a robust change in the concentration of oxyHb takes place over the temporo-parietal region of the LH. Interestingly, the concentration of oxyHB is relatively stable both in the BW and in the Silence conditions. Forward speech gives rise to a significant increase in oxyHb over the LH. No significant change is observed when BW speech is used. While the energy is identical in FW and BW, and their spectral properties are mirror images of each other, only FW gives rise to a significant increase of deoxyHB over the LH. Figure 8.1 illustrates these results.

These results suggest that the brain of the newborn infant responds differently to natural and backward speech. To understand the singularity of this result, the reader has to remember that monolingual adults who are tested with similar materials as the infants are sometimes tricked to believe that both FW and BW are sentences in some foreign languages. Interestingly, if they are asked to rate which one sounds more "natural," they tend to choose forward speech. The BW and FW utterances are indeed very similar but they differ at the suprasegmental level. FW and BW speech differ in terms of the development of their timing patterns. Indeed, final lengthening appears to be a universal property of natural language. Thus, BW utterances have initial lengthening. In addition, some segments (stops i.e., [p], [t], [k], [b], [d], and [g] and affricates, like [ts] or [dz]) become very different when played BW. The vocal tract cannot produce BW speech. Since infants cannot produce FW speech either, they might ignore the contrast between the BW and FW conditions (see Liberman & Mattingly, 1985). Since the neonate's brain responds in a different way to FW and BW utterances, we suggest that babies, in some sense of "know," know the difference between utterances that can and cannot be articulated by humans. We might tentatively attribute this result to the specialization of certain cortical areas of the neonate's brain for speech. Humans might have, like many other vertebrates, specialized effectors and receptors for a species-specific vocalization, which in our case is speech. This possibility needs to be studied in greater detail.

The above results have to be evaluated with care. Results from the work we have carried out with nonhuman organisms show that they display a behavior similar to that of infants, when confronted with FW and BW speech. In a series of studies comparing the newborn infant and the tamarin monkey behavioral responses, Ramus *et al.* (2000) showed that like infants, tamarins discriminate two languages when the utterances are played forward but fail to do so when the utterances are played backward. Tamarins will never develop speech, yet they notice the change from FW to BW speech. This ought to temper any desire to conclude that the above results are based on a species-specific system to process natural speech. They may also suggest that the specialization may be more basic, that is, not for speech as such but for sounds produced by vocal tracts that emit air through a narrow passage. Higher vertebrates produce sounds in this way.

In an attempt to replicate and expand the above experiment, a new device was used to measure simultaneously activation over 12 positions on the RH and 12 on the LH (see Peña, Maki, Dehaene-Lambertz, Bouquet, Koizumi & Mehler, 2003). The design of the experiment was otherwise identical to the one described above. The outcome shows that the overall pattern of activation mimics that already observed with the more primitive device. Indeed, we found that the infant's brain is activated by acoustic stimuli, regardless of whether these are FW or BW speech as compared to no stimulation. However, we also found that the total HB response to FW is larger on the LH than on similar areas of the RH. This is not the case for BW. Indeed, for BW speech, the total HB response is comparable on the RH and the LH. These results suggest that normal speech is differently processed to a very well-matched control, namely BW speech.

Obviously, the advent of imaging studies with neonates will permit new and more precise investigations to establish whether the specialization for speech is really present at birth or whether there is activation for streams of sounds that can be produced by a vertebrate's vocal tract. We believe that these kinds of study will set in motion new investigations that will clarify the validity of many of our current views. In the meantime, these studies have shed some light into complex issues that were hard to study with more traditional behavioral methods.

8.4 Neonates use rhythm to tune into language

Rhythm is a percept that relates to the relative duration of constituents in a sequence. What are the elements responsible for rhythm in language? Three constituents have been proposed to be roughly isochronous in different languages, thus giving rise to rhythm: syllables, feet, and morae (see Abercrombie, 1967; Ladefoged, 1975; Pike, 1945). Syllables have independently been construed as a basic constituent or atom in speech production and comprehension (see Cutler *et al.*, 1983; Levelt, 1989; Mehler, 1981). Infants begin to produce syllables several months after birth, with the onset of babbling. However, the infant may process syllables before he/she produces them. If so, we ought to find precursors illustrating that neonates process syllables in linguistic-like ways.[5] Bertoncini (1981) explored this issue using the nonnutritive sucking technique showing that very young infants distinguish a pair of syllables that differ only in the serial order of their constituents segments, for example, *PAT* and *TAP*. The infants, however, fail to distinguish a pair of items derived from the previous ones by replacing the vowel [a] by the consonant [s]. This renders the items *TSP* and *PST*, impossible syllables. To understand the infant's failure to distinguish this pair, in a control experiment, infants were presented with the same items but surrounded by a vocalic context. When the same sequences are presented in a syllabic context, as when they are surrounded by a vowel (as in *UPSTU* and *UTSPU*), the infant's discrimination ability is restored. This experiment suggests that the infant makes distinctions in linguistic-like contexts that are neglected in other acoustic contexts.

As we mentioned in Note 6, some languages (e.g., Croatian, some varieties of Berber, etc.) allow specific consonants to occupy the syllabic nuclear position. For instance, in Croatian, *Trieste*, the Italian city, is named *Trst* where [r] is the nucleus. This is not an exceptional case in the language. Indeed, the word for "finger" is *prst* and the word for "pitcher" is *vrč*. Why then were the results reported in the previous experiment obtained? Why did the infants neglect to treat *PST* and *TSP* as syllables? Maybe we tested infants who were already rather old, i.e., 2 months, and thus had already considerable exposure to the surrounding language. Since they are all raised in a French environment, it is possible that the stimuli were already considered extraneous to their language and thus their differences neglected. Alternatively, PST and TSP are impossible syllables in any language. To the best of our knowledge, in fact, there is no language that allows [s] as a syllabic nucleus. We are currently exploring means to choose between these two alternative explanations. We predict that infants have no difficulties in distinguishing pairs in which [r] or [l] figure as nuclei (e.g. [prt] vs. [trp] or [plt] vs. [tlp]) since such syllables occur in a few languages but that they will have difficulty distinguishing sequences in which the nuclear position is occupied by [s] or [f] (e.g. [pst] vs. [tsp] or [pft] vs. [tfp]). To insure that the infant has not become familiar with the syllable repertoire in the surrounding language, we are testing neonates in their first week of life.

That infants are attending to speech using syllabic units has also been claimed by Bijeljac-Babic *et al.* (1993). These authors showed that infants distinguish lists of bi-syllabic items from a list of tri-syllabic ones. They used CVCV items (e.g., *maki, nepo, suta, jaco*) and CVCVCV items (e.g., *makine, posuta, jacoli*). This result is observed regardless of whether the items differ or are matched for duration. Indeed, some of the original items were compressed and others expanded to match the mean durations of the two lists. Infants discriminated the lists equally well, suggesting that it is the number of syllables or just the number of vowels in the items that counts for their representation. We have had to focus on syllables rather than feet or morae because few studies have explored whether neonates represent these units. Below we are going to explain why we believe that syllables, or possibly vowels, play such an important role during the early steps of language acquisition.

The results described above fit well together with recent evidence showing that neonates are born with remarkable abilities to learn language. For instance, in the last decade numerous studies have uncovered the exceptional abilities of babies to process the prosodic features of utterances (see Mehler *et al.*, 1988; Moon *et al.*, 1993). Indeed, for many pairs of languages, infants tend to notice when a speaker switches from one language to another. What is the actual cue that allows infants to detect this switch? The essential property appears to be linguistic rhythm, defined as the proportion that vowels occupy in the utterances of a language (see Ramus *et al.*, 1999). If two languages have different rhythms (an important change in %V), the baby will detect a switch from one language to the other. If languages have similar rhythms, as for instance, English and Dutch or Spanish and Italian, very young infants will fail to react to a switch (see Nazzi *et al.*, 1998).

The variability of the inter-vocalic interval (i.e., ΔC, the standard deviation of the intervocalic intervals) also plays an important role in explaining the infants' behavior. In fact, ΔC in conjunction with %V provides an excellent measure of language rhythm that fits well with the intuitive classification of languages that phonologists have provided. Indeed, their claim is that there are basically three kinds of rhythm depending on which of three possible units maintains isochrony in the speech stream: stress-timed rhythm, syllable-timed rhythm and mora-timed rhythm (see Abercrombie, 1967; Ladefoged, 1975; Pike, 1945). However, once exact measures were carried out, contrary to many an expectation, isochronous units were not found (see Dauer, 1983; Manrique & Signorini, 1983; but see Port *et al.*, 1987). This does not mean, as one might have argued, that the classification linguists proposed on the basis of their intuitions has to be dismissed. Rather, Ramus *et al.*'s definition of rhythm on the basis of ΔC and %V divides languages exactly into those three intuitive classes, as shown in Figure 8.2.

A language with a high %V and a small ΔC (like Japanese or Hawaiian) is likely to have a small syllabic repertoire. Mostly, such languages allow only CVs, and Vs giving rise to the typical rhythm of the mora-class. Moreover, intervocalic intervals cannot be very variable since consonant clusters are avoided and codas are in general disallowed. In Japanese, for instance, codas generally contain /n/ (as in the word *Honda*).[6] Romance languages, as depicted in Figure 8.2, have a smaller value of %V because their syllabic repertoires are larger. Indeed, these languages allow both onsets and codas. Moreover, onsets may contain consonant clusters and occasionally also codas contain more than one consonant (e.g., *prêt*, *sparo*, *tact*, *parc*, etc.). However, fewer syllable types are allowed in Romance languages than in stress-timed languages as Dutch and English. Indeed, while in

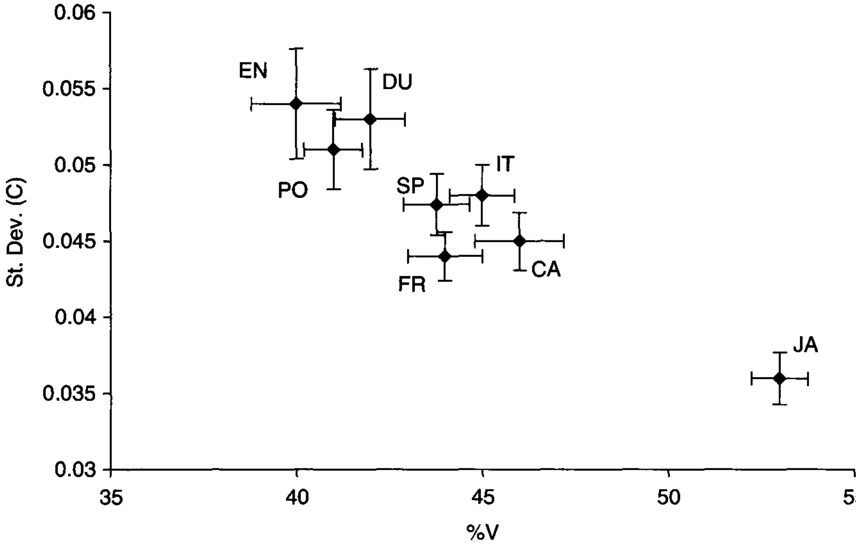

Figure 8.2 %V is the mean proportion of the utterances in a language that is occupied by vowels and ΔC or St. Dev. (C) is the standard deviation of the consonantal intervals. The plot incorporates eight languages spoken by four female speakers. Each speaker utters 20 sentences (each language is represented by 80 utterances). The distribution of the languages is compatible with the notion that they can be grouped into three classes as predicted by phonological intuitions (from Ramus *et al.*, 1999).

Romance languages the typical syllabic repertoire ranges from 6 to 8 syllables, Germanic languages have over 16 syllable types. This conception of rhythm relates to Dauer (1983) and also Nespor (1990) who claim that linguistic rhythm is a side effect of the syllabic repertoire that languages instantiate. Languages such as Japanese have a very restricted syllable repertoire, and thus a relatively high proportion of utterances is taken up by vowels. In contrast, languages with a large number of syllable types, thus many consonant clusters, tend to have a smaller proportion of utterances taken up by vowels. Interestingly, one could conclude that after a larger number of languages is included in Figure 8.2, it might turn out that some more classes or even a continuum is obtained rather than the clustering of languages into the few classes that we now observe. However, if the notion of rhythm is really related to the claim according to which the number of syllable types is what gives rise to the intuitive notion of linguistic rhythm, things will go in favor of a clustering. Indeed, the syllable repertoires come in groups. Up until now, we have languages that have 2 or 3 syllable types (Hawaiian, Japanese, etc.), 6 to 10 syllable types (Spanish, Greek, Italian, etc.) and languages that have 16 or more (English, Dutch, etc.) (see Nespor, 1990). Future scrutiny with a larger set of languages will determine whether the notion that languages fall into a restricted number of classes is born out or not; and if so, how many classes there are.

We are willing to defend the conjecture that languages cluster into a few classes, because rhythm, as defined by Ramus *et al.* (1999), is sufficient to explain the available behavioral results. Indeed, Ramus *et al.* (1999) simulated the ability to discriminate switches from one language to another in infants and adults. He showed that %V is sufficient to account for all the empirical findings involving neonates. This outcome sustains our resolve to pursue this line of investigation. Indeed, it is unlikely that linguistic rhythm would play such an important role in determining the neonate's behavior without having any further influence on how language is learned.

The first adjustment the neonate makes to the surrounding language concerns rhythm. The processing of linguistic rhythm appears to change over the first 2 months of life. Mehler *et al.* (1988) remarked that while American 2-month-olds fail to discriminate Russian from French, 1-week-old French infants successfully discriminate not only Russian from French but also English from Italian. The authors argued that by 2 months of age infants have encoded some properties of their native language and stop discriminating between two unfamiliar rhythms. Such a bias may explain the observed failure to discriminate a switch between two "unknown" languages. Christophe and Morton (1998) further investigated this same issue testing 2-month-old British infants. They found that the infants were able to discriminate a switch between English and Japanese but not a switch between French and Japanese. Presumably, the former pair of languages is discriminated because it entails one familiar and one novel rhythm. The second switch yields no response because neither language has a familiar rhythm. To buttress their interpretation, Christophe and Morton (1998) also tested the behavior of these same British infants with Dutch. First, they corroborated their prediction that these infants would fail to discriminate Dutch from English, because the two languages have a similar rhythm. Next, they showed that the infants discriminate Dutch from Japanese, two foreign languages for these infants. In fact, while Dutch differs from English, their rhythm is similar, and thus, although Dutch is not their native language it still catches the infants' attention.

Pure behavioral research may be insufficient to ground the above explanations. We hope, however, that adequate brain-imaging methods, used as indicators of processing, could provide more information to decide whether learning and development of language requires a passage through an attention-drawing device based on rhythm.

Why are infants interested in rhythm even before the segments of the utterances capture their curiosity?[7] What information does linguistic rhythm provide to render it so relevant for language acquisition? We have followed two procedures to answer these questions. First, we have tried to gather data using optical topography (see above pp. 214–217), to pursue the exploration of language processing in the neonate, as described above. Second, we have explored the potential role of rhythm in other areas of language acquisition. Specifically, we asked whether rhythm may play a role in the setting of syntactic parameters, and also whether it might be exploited in segmentation, as described in the following sections.

8.5 Segmenting the speech stream

Ramus *et al.* (1999) (see Section 3.4) conjectured that language rhythm provides the infant information about the richness of the syllabic repertoire of the language (cf. Dauer, 1983; Nespor, 1990).

For the sake of the argument, we assume that the infant gains this type of information from the rhythmic properties in the signal. What would then be the use of such information for the language-learning infant? What profit does the baby draw by knowing that the number of syllable types is 4, 6, or 16? Will such information facilitate perception of speech? Or will such information be essential to master the production routines or elementary speech acts? We cannot answer these questions in detail. However, there is yet no reason to believe that knowing the size of the syllabic repertoire facilitates perception of speech. Is there evidence that a learner performs better when he/she has prior knowledge of the number of types or items in the set to be learned? We can give an indirect answer by looking at lexical acquisition. Surely, infants learn the lexicon without ever knowing or caring whether they have to master 4000 or 40 000 words. Why would knowledge of the number of syllable types be useful compared to learning the syllables in the language much as one learns words? There is no ready answer to this question, which does not mean that in the future an answer will not be forthcoming. However, there is an explanation for the infant's precocious interest in rhythm. Rhythmic information may constrain lexical acquisition. Indeed, the size of the syllabic repertoire is inversely correlated with the mean length of words. Hence, gaining information about rhythm may provide through an indirect route a bias as to the average size of the lexical items in the language of exposure (Mehler & Nespor, 2004).

When listening to connected speech, the baby has to break up the input into constituent-like words. However, it is well known that speech signals do not afford reliable acoustic cues about the beginning and the end of words. The most naïve psycholinguistic explanation of parsing is to postulate that there are gaps between words but in fact there are none. Prosodic cues may signal the end of a word that is found at the right edge of larger constituents, such as phonological phrases or intonational phrases, but not every word of the speech stream. In fact, even when gaps are found, they are as likely to fall within words as between words, for example because the release of a voiceless stop is preceded by a constriction that very much looks like a pause.

How can rhythm help segmenting the continuous speech stream? Mehler and Nespor (2004) have proposed that infants who listen to a language with a %V that is higher than 50%, like in "mora-timed" languages, will tend to parse signals looking for long constituents while infants who listen to a language whose %V is below 40% will tend to search for far shorter units (see p. 00 for details). This follows from the fact that the syllabic repertoire in, for example, Japanese is very limited,[8] which entails that monosyllables will be rare and long words will be very frequent, unless speakers are willing to put up with polysemy to such an extent as to threaten communication. However, languages are designed to favor rather than to hinder communication. Hence, words turn out to be long in Japanese as well as in any other language with a restricted syllabic repertoire. In contrast, languages such as Dutch or English, which have a very rich syllabic repertoire (%V close to 45%), allow for a large number of different syllables; hence, without increasing ambiguity one can imagine that among the first 1000 words in the language many will be monosyllables (nearly 600 out of 1000). Languages like Italian, Spanish, or Catalan, whose %V lies between that of Japanese and that of English, also have an intermediate number of syllable types. As expected, the length of the most common words falls between two and three syllables.

Assuming that rhythmic properties are important during language acquisition and, furthermore, that very young infants extract the characteristic rhythm of the language of exposure, it would be nice to know the computational processes that allow such an extraction to take place. Unfortunately, at this time, we have no concrete results that would allow us to explain how these computations are performed. Hopefully, future studies will clarify whether the auditory system is organized to extract rapidly and efficiently the rhythmic properties of stream of speech, and/or whether we are born to be powerful statistical machines so that small differences in rhythm between classes of languages can be ascertained. Independent of how the properties that characterize the rhythmic classes are identified, our conjecture is that the trigger that biases the infant to expect words of a certain length is determined by rhythm. Once rhythm has set or fixed this bias, one may find that infants segment speech, relying on other mechanisms. For example, the statistical computations that Saffran and her colleagues have invoked (see below) may be an excellent tool to segment streams of speech into constituents. However, it is possible that the rhythm in the stream will bias the learner to go for longer or shorter items depending on the language they are learning.

Saffran *et al.* (1996) and Morgan and Saffran (1995) have revived the view that statistical information plays a central role in language acquisition. Indeed, information theorists (Miller, 1951) had already postulated that the statistical properties of language could help process signals and acquire parts of language. Connectionism has also highlighted the importance of statistics for language learning. They have even gone as far as viewing the language learner as a powerful statistic machine. Without going as far as those investigators have gone, we recognize that the advantage of statistics is that it can be universally applied to unknown languages, and thus pre-linguistic infants may also exploit it.

Saffran *et al.* (1996) have shown that adults and 9-month-old infants confronted with unfamiliar monotonous artificial speech streams tend to infer word boundaries through the statistical regularities in the signal. A word boundary is postulated in positions where the transitional probability (hereafter TP) drops between one syllable and the next.[9] Participants familiarized with a monotonic stream of artificial speech recognize tri-syllabic items delimited by dips in TP. As an example, imagine that *puliko* and *meluti* are items with high TPs between the constituent syllables. If Ss are asked which of *puliko* or *likome* (where *liko* are the last two syllables of the first word and *me* the first syllable of the second word) is more familiar, they tend to select the first well above chance. Among a large number of investigations that have validated Saffran *et al.*'s findings, we have found that, by and large, French and Italian adult speakers perform as the English speakers of the original experiment.[10]

Let us summarize what we have tried to suggest this far. We have noticed that linguistic rhythm can be captured as suggested by Ramus *et al.* (1999) by measuring the amount of time/utterance occupied by vowels and by the variability of the intervocalic intervals. This proposal presupposes that our processing system makes a categorical distinction between consonants and vowels. In the following section, we expand on the notion that there is a basic categorical distinction between Vs and Cs, and we go on to propose a view of language acquisition based on the consequences of this divide.

8.6 Rhythm, signals, and triggers

Developmental psycholinguists and students of adult language perception and production considered the possibility that different phonological units are highlighted depending on the rhythmic class to which a language belongs, as described above. More recently, linguists and psycholinguists started exploring whether the different phrasal phonological properties related to syntax can guide the infant in the setting of parameters that are essential to acquire language. We are presently exploring to what extent linguistic rhythm can help the learner discover some of the nonuniversal properties of syntax. It is in this research area that the investigation of the syntax–prosody interaction might offer a link between an exclusively syntactic approach to PS and the cognitive neuroscience approach, which concentrates on the perception and production of speech.

The acquisition of some aspects of language is facilitated by the statistical properties encoded in the speech signal (see p. 224). For most classical association accounts of acquisition, the more a property is transparently encoded in the signal, the easier it will be to learn, regardless of the domain - including language. Such theories assume that the signals are rich enough to inscribe structure in the head of the learner. No innate knowledge is postulated over and beyond the ability to associate signals. In contrast to classical learning, linguists have argued that in order to learn to speak a language one must learn grammar. For this to happen, they argue, innate knowledge has to be postulated because general learning mechanisms are not sufficient to allow the infant to acquire grammar directly from the signal. The nature of this knowledge is roughly spelled out in the PS theory. We believe that this is the richest account of language acquisition we are aware of because it relates universal principles to aspects of grammar that are language specific. As we stated before, this theory might be correct or not. However, it is the only theory that can be explored in sufficient detail as to allow its dismissal if it does not mesh well with observation.

Our proposal is to integrate PS with a general theory of learning. While it is commonly taken for granted that general learning mechanisms play a role in the acquisition of the lexicon (Bloom, 2000), their role in the actual setting of the parameters has not been sufficiently explored. In fact, while signals might give a cue to the value of a certain parameter, general learning mechanisms might play a role in establishing the validity of such a cue for the language of exposure. For instance, in order to decide whether in a language complements precede or follow their head, it is necessary to establish whether the main prominence of its phonological phrases is rightmost or leftmost, as we will see below. Within a language, syntactic phrases, by and large, are of one type or another: that is they are either Head–Complement (HC) or Complement–Head (CH). There are languages, however, in which a specific phrase might have a word order different from the standard word order of the language. Since the pre-lexical infant ignores whether this exception weakens the relation of prominence with the underlying parameter, it needs a mechanism to cope with the presence of this confusing information. In all likelihood, statistical computations allow the infant to discover and validate the most frequent phonological pattern that can then be used as a cue to the underlying syntax (see Nespor *et al.*, 1996). Even if such exceptional patterns did not exist in a language, the need for statistics remains plausible. Indeed, even an infant that is exposed to a regular language (as to the HC order) might occasionally hear irregular patterns, for example foreign locutions or speech errors. In this case, the frequency distribution difference between the occasional and the habitual patterns will allow the infant to converge to the adequate setting.

Let us focus in more detail on the case of the HC parameter. This is a central parameter for learning the syntax of one's language. Indeed, in the great majority of languages, the setting of this parameter simultaneously specifies the relative order of heads and complements and of main clauses with respect to subordinate clauses. That children start the two-word stage without making mistakes in the word order suggests that this parameter is set precociously (see Bloom, 1970; and also Meisel, 1992). In addition, before that, they react differently to the appropriate, as compared to the wrong, word order (Hirsh-Pasek & Golinkoff, 1996). These facts suggest that children must set this parameter quite early in life.

Given our viewpoint, it would be quite desirable to imagine a scenario in which the infant finds ways and means to set basic parameters prior or at least independently of the segmentation of the speech stream into words. If the child sets parameters before learning the meaning of words, prosodic bootstrapping would become immune to the paradox pointed out by Mazuka (1996). She observes that to understand the word order of, say, heads and complements in the language of exposure, an infant must first recognize which is the head and which is the complement. But once the infant has learned to recognize in a pair of words which one functions as head and which as complement, it already knows how they are ordered. If you know how they are ordered, the parameter becomes pointless. Without syntactic knowledge, word meaning cannot be learned and without meaning, syntax cannot be acquired either.

How can a child overcome this quandary and get information about word order just by listening to the signal? What is there in the speech stream that might give a cue to the value of this parameter? Rhythm, in language as in music, is hierarchical in nature (see Liberman & Prince, 1977; Selkirk, 1984). We have seen above that at the basic level, rhythm can be defined on the basis of %V and ΔC. At higher levels, the relative prominence of certain syllables (or the vowels that form their nucleus) with respect to other syllables reflects some aspects of syntax. In particular, in the phonological phrase,[11] rightmost main prominence is characteristic of head-complement languages, like English, Italian, or Croatian while leftmost main prominence characterizes complement-head languages, like Turkish, Japanese, or Basque (Nespor & Vogel, 1986). A speech stream is thus an alternation of words in either weak–strong or strong–weak chunks. Suppose that this correlation between the location of main prominence within phonological phrases and the value of the HC parameter is indeed universal. Then we can assume that by hearing either a weak–strong or a strong–weak pattern, an infant becomes biased to set the parameter to the correct value for the language of exposure. The advantage of such a direct connection between signal and syntax (see Morgan & Demuth, 1996) is that the only prerequisite is that infants hear the relevant alternation. To see whether this is the case, Christophe *et al.* (1997) and Christophe *et al.* (2003) carried out a discrimination task using resynthetized utterances drawn from French and Turkish sentences. These languages have similar syllabic structures and word final stress but they differ in the locus of the main prominence in the phonological phrase, an aspect that is crucial for us.[12] The experiment used delexicalized sentences pronounced by the same voice.[13] Infants 6- to 12-weeks-old discriminate French from Turkish. It is concluded that infants discriminate the two languages only on the basis of the different location of the main prominence. Knowing that infants discriminate these two types of rhythmic patterns opens a new direction of research to assess whether infants actually use this information to set the relevant syntactic parameter.

8.7 The C/V distinction and language acquisition

Why does language need to have both vowels and consonants? According to Plato, rhythm is "order in movement." But why, at one level of the rhythmic architecture, is the order established by the alternation of vowels and consonants? Why do all languages have both Cs and Vs? Possibly, as phoneticians and acousticians argue (see Stevens, 2000), this design structure has functional properties that are essential for communication. Indeed, vowels have considerable energy, allowing them to carry the signal, while consonants are modulations that allow increasing the number of messages with different meaning that can be transmitted. Even if one believes that this explanation is correct, it may not be the only one of the reasons why languages necessarily include both vowels and consonants.

Nespor *et al.* (2003) has proposed that vowels and consonants, because of their different phonetic and phonological properties, play a different functional role in language acquisition and language perception. The main role of consonants is to be intimately involved with lexical structure, while that of vowels is to be linked to grammatical structures.

The lexicon allows the identification of thousands of lemmas, while grammar organizes the lexical items in a regular system. There is abundant evidence that consonants are more distinctive than vowels. For instance, cross-linguistically there is a clear tendency for Cs to outnumber Vs: the segmental system most frequent in the languages of the world has 5 vowels and around 20 consonants. But languages with just 3 vowels are also attested and historical linguists working on common ancestors of different languages have posited two or even one vowel for proto-Indo-European.

A widespread phenomenon in the languages of the world is to reduce vowels in unstressed positions. Languages like English, in which unstressed vowels are centralized to schwa, thereby losing their distinctive power, represent an extreme case. No comparable phenomenon affects consonants. The pronunciation of Cs is also less variable (thus more distinctive) than that of Vs. Prosody is responsible for the variability of vowels within a system: both rhythmic and intonational information (be it grammatical or emotional) is by and large carried by vowels. Acoustic-phonetic studies have documented that while the production of vowels is rather variable, consonants are more stable. Moreover, experimental studies have shown that while consonants are perceived categorically, vowels are not (Kuhl *et al.*, 1992; Werker *et al.*, 1984). These different reasons for the variability of vowels, of course, make them less distinctive. Evidence for the distinctive role of consonants is also attested by the existence of languages (e.g., Semitic languages) in which lexical roots are composed uniquely by consonants. To the best of our knowledge, there is no language in which lexical roots are composed just of vowels.

The above noted asymmetry between Vs and Cs in linguistic systems is reflected in language acquisition. The first adjustments infants make to the maternal language are related to vowels rather than to consonants. Indeed, several pieces of evidence can be advanced to buttress this assertion. In a study, Bertoncini *et al.* (1988) showed that very young infants presented with four syllables in random order during familiarization react when a new syllable is introduced, provided that it differs from the others by at least its vowel. If the new syllable differs from the other syllables only by the consonant, its addition will be neglected.[14] However, 2-month-olds show a response to both, that is whether one adds a syllable that differs from a member of the habituation set by its vowel or by its consonant. We must remember, however, that the above results are not due to limitations in discrimination ability but rather to the way in which the stimuli are represented.[15] We can conclude that the first representation privileges vowels but that by 2 months of age vowels and consonants are sufficiently well encoded as to yield a similar phonological representation. In fact, by 6 months of age infants respond preferentially to the vowels of their native language.[17] In contrast, Werker and her colleagues have shown that consonant contrasts that are discriminated before 8 months are neglected a few months later if they are not used in the maternal language (Werker & Tees, 1984); that is, when the infant goes from phonetic to phonological representations, vowels seem to be adjusted to the native values before consonants. This observation is yet another indication that vowels and consonants are categorically distinct from the onset of language acquisition. Our suggestion is that these two categories have a different function in language and in its acquisition.

As we mentioned above (see p. 227), vowels and consonants, even when they are equally informative from a statistical point of view, are not exploited in similar ways. Newport & Aslin (2004) used a stream of synthetic speech consisting of CV syllables of equal pitch and duration in which the vowels change constantly and "words" are characterized only by high TPs between the consonants. Participants successfully segment such a stream.[18] We replicated this robust finding with Italian and French-speaking subjects (Bonatti *et al.*, 2005). In a similar experiment in which the statistical dependences were carried by vowels while the intervening consonants vary, the participant in our experiment failed to segment the stream into constituent "words." Thus, a pre-lexical infant (or an adult listening to an unknown language) identifies word candidates on the basis of TP dips between either syllables or consonants, but not between vowels. However, see FN ... Why should this be so? As pointed out above, consonants change little when the word is pronounced in different emotional or emphatic contexts while vowels change a lot. Moreover, a great number of languages introduce changes in the vowels that compose a group of morphologically related words, that is *foot-feet* in English, and more conspicuously, in Arabic: *kitab* "book," *kutub* "books," *akteb* "to write." In brief, consonants rather than vowels are mainly geared to insure lexical functions. Vowels, however, have an important role when one attempts to establish grammatical properties. We argued above that the rhythmic class of the first language of exposure is identified on the basis of the proportion of time taken up by vowels. Identifying the rhythm, we argued, provides information about the syllable repertoires, that is a part of the phonology. Moreover, it gives information about the mean length of words in the language. Also, a piece of information carried by vowels relates to the location of the main prominence within the phonological phrase. As was argued above, prominence is related to a basic syntactic parameter.

8.8 Conclusion

In this chapter, we have argued that both innate linguistic structure and general learning mechanisms are essential to our understanding of the acquisition of natural language. Linguists have paid a lot of attention to universal principles or constraints that delimit the nature of our endowment for language. Psychologists, in contrast, have focused on how the child acquires the language of exposure, without being concerned with the biological underpinnings of this achievement. After scrutinizing the limitations of both positions, we have pleaded for an integration of the two approaches to the study of language acquisition. Currently, there is a growing consensus that biologically realistic models have to be elaborated in order to begin understanding the uniqueness of the human mind and in particular of language.

In our research, we have highlighted the importance of exploring how signals relate to the fixation of parameters. We have tried to demonstrate that signals often contain information that is related to unsuspected properties of the computational system. We laid out a proposal of how rhythm can guide the learner toward the basic properties of the language's phonology and syntax. We have also argued that basic phonological categories, namely vowels and consonants, play different computational roles during language acquisition. These categories play distinctive roles across languages and appear to be sufficiently general for us to conjecture that they are a part of the species' endowment.

Another aspect that we highlighted concerns the attested acoustic capacity of vertebrates to discriminate and learn phonetic distinctions (see Kluender *et al.* 1998; Ramus *et al.*, 2000, etc.). They also have the ability to extract and use the statistical properties of the stimulating sequences in order to analyze and parse them into constituents (M. Hauser, personal communication). These results suggest that humans and other higher vertebrates can process signals much in the same way. However, the fact remains that only humans, and no other animals, acquire the language spoken in the surrounds. Moreover, simple exposure is all that is needed for the learning process to be activated. Thus, we must search for the prerequisites of language acquisition in the knowledge inscribed in our endowment.

The fact that cues contained in the speech stream directly signal nonuniversal syntactic properties of language makes it clear that to understand how the infant attains knowledge of syntax precociously and in an effortless fashion, attention must be paid to the very cues that the signals provide. How can this argument be sustained when we have just acknowledged that human and nonhuman vertebrates process acoustic signals in a similar fashion? Because, a theory of language acquisition requires not only an understanding of signal processing abilities but also of how these cues affect the innate linguistic endowment. The nature of the language endowment, once precisely established, will guide us toward an understanding of the biological foundation of language, and thus will clarify why we diverge so significantly from other primates. This in turn will hopefully lead us to formulate a testable hypothesis about the origin and evolution of natural language.

Notes

1. To illustrate this, consider a child who hears mostly sentences with a Verb–Object order. The child, putatively, obtains automatically information from the linguistic input to set the relevant word-order parameter. If this were so, it would constitute a great asset, since fixing the word-order parameter may greatly facilitate the acquisition of grammar and also the acquisition of the lexicon. Likewise, the child exposed to a language that can have sentences without an overt subject, for example Italian ("piove," "mangiano arance," etc.), or to a language whose sentences require overt mention of subjects, for example English ("it rains," "they eat oranges"), supposedly gets information from the linguistic input to set the relevant parameter.
2. See Hebb, D. O. (1949). *Organization of Behavior.* New York: Wiley.
3. To establish that the FFA is an area that is specifically triggered by faces, Kanwisher and also others had to test many other stimuli and conditions. Even so, Gauthier and her collaborators have challenged the existence of the FFA showing that this area is also activated by other sets whose members belong to a categorized ensemble even though they are not faces. Moreover, Gauthier and her colleagues showed that when Ss learn a new set before the experiments, its members then activate the FFA activation. Gauthier argued that her studies show that the FFA is not uniquely a structure devoted to face processing. Without denying the validity of Gauthier's results, Kanwisher still thinks that the FFA is a *bona fide* face area. We think that although we understand much better the FFA than the Belin's voice, we still have to be very careful before we accept the proposed locus as a voice-specific area. A fortiori we need equal parsimony before we admit that we do have a specific voice-processing area. Future research will clarify this issue.
4. This device uses near-infrared light to evaluate how many photons are absorbed in a part of the brain following stimulation. The device is light and non-invasive. In this sense, it is comparable to most Evoked Response Potential devices currently in use. The difference is that like fMRI it estimates the vascular response in a given area of the cortex. Another difference is that like fMRI its time resolution is poorer than that of ERP. Our device uses bundles of fiber optics that are applied to the infants' head. These light bundles contain a fiber that delivers near-infrared light of two wavelengths. The other fiber, which is placed 3 cm away from the irradiating one, is a light-collector fiber. One of the wavelengths is absorbed by oxyHb while the other is absorbed by deoxyHb. When one measures the changes in emerging light for each wavelength, it is possible to estimate precisely the functional organization of the underlying cortical areas.
5. A universal property of syllables is that they have an obligatory *nucleus* optionally preceded by an *onset* and followed by a *coda*. While onset and coda are occupied by consonants (C), the nucleus is generally occupied by a vowel (V). In some languages, the nucleus can be occupied by a sonorant consonant (as [m], [n],

[l], and [r], in particular [r]). Thus, a syllable may not contain more than one vowel. CV is the optimal syllable, that is the onset tends to be present and the coda absent. All natural languages have CV syllables. There is a hierarchy of increasing complexity in the inclusion of syllable types in a given language. Thus, a language that has V will also have CV, but not vice versa. A language that has V, instead, does not necessarily have VC. That is, in some languages all syllables end in a vowel. Similarly, a language that has CVC will also have a CV in its repertoire. A language that includes a CCV in its repertoire will have CV and a language that includes CVCC also has CVC. The prediction then is that while CVC is a well-formed potential syllable in many languages, CCC is not, in particular if none of the consonants is sonorant.

6. Or geminates as in the word *Sapporo*.
7. Werker and Tees (1983) were the first to point out that the first adjustment to the segmental repertoire of the language of exposure becomes apparent at the end of the first year of life.
8. Syllable types in Japanese are CV and V. Coda consonants are limited to be either an [N] or a geminate consonant shared with the following syllable.
9. Saffran, *et al.* use streams that consist of artificial CV syllables that are assembled without leaving a pause between one another. All syllables have the same duration, loudness, and pitch. TPs between adjacent syllables (in any trisyllable) range from 0.25 to 1.00. The last syllable of an item and the first syllable of the next one have TPs ranging from 0.05 to 0.60.
10. One divergence between the results reported by the Rochester group and our own concerns the computation of TPs on the consonantal and vocalic tiers. Apparently, native English speakers can use both tiers to calculate TPs (see Section 8.7). Our own Ss, regardless of whether they are native French or native Italian speakers, can only use the consonantal tier, see p. 27 for more details.
11. The phonological phrase is a constituent of the phonological hierarchy that includes the head of a phrase and all its function words. It also includes some complements and modifiers under specific syntactic conditions as well as conditions concerned with weight (Nespor & Vogel, 1986).
12. The effect of the resynthesis is that all segmental differences are eliminated.
13. Sentences were synthetized using Dutch diphones with the same voice.
14. Two kinds of habituation were used, [bi], [si], [li], and [mi] or [bo], [bae], [ba], and [bo]. The introduction of [bu] causes the neonate to react to the modification regardless of the habituation. The introduction of [di] after the neonate is habituated with the first set of syllables neglected and so is the introduction of [da] after habituation with the second set.
15. In discrimination experiments, one evaluates whether infants react when a repeated syllable suddenly changes. In the present study, one evaluates whether the infant reacts when a set of four repeated syllables suddenly includes a novel syllable. In this case, one tests the details with which the initial set of syllables was represented

rather than a simple discrimination.

16. American infants respond preferentially to American vowels as compared to Swedish vowels while Swedish infants respond preferentially to Swedish vowels compared with English ones (see Kuhl, P. K., Williams, K. A., *et al.*, 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, **255**, 606–8).

17. Thus, if a word has the syllables C-, and C′-, C″ with the consonants that predict the next one exactly, regardless of the vowels that appear between them, it will be preferred to a part word like C″-, C*-, and C**, where the stars illustrate that the two last syllables come from another "word." Of course, words have no probability dip between the consonants but part words enclose a TP dip between C″ and C*.

References

Abercrombie, D. (1967). *Elements of General Phonetics*. Chicago: Aldine.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, **403**(20 January), 309–12.

Bertelson, P. (1982). Lateral differences in normal man and lateralization of brain function. *International Journal of Psychology*, **17**, 173–210.

Bertoncini, J. (1981). Syllables as units in infant speech perception. *Mehler, Journal of Infant Behavior and Development*, **4**, 247–60.

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

Bertoncini, J., Morais, J., Bijeljac-Babic, R., McAdams, S., Peretz, I. and Mehler, J. (1989). Dichotic perception and laterality in neonates. *Brain and Cognition*, **37**, 591–605.

Best, C. T., Hoffman, H. and Glanville, B. B. (1982). Development of infant ear asymmetries for speech and music. *Perception and Psychophysics*, **31**, 75–85.

Bijeljac-Babic, R., Bertoncini, J. and Mehler, J. (1993). How do four-day-old infants categorize multisyllabic utterances? *Developmental Psychology*, **29**, 711–21.

Bloom, L. (1970). *Language Development: Form and Function in Emerging Grammars*. Cambridge, MA: MIT Press.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.

Bonatti, L. L., Peña, M., Nespor, M. and Mehler, J. (2005). Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychological Science*, **16**, 451–9.

Bryden, M. P. and Allard, F. A. (1981). Do auditory perceptual asymmetries develop? *Cortex*, **17**, 313–18.

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

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.

Chomsky, N. (1986). *Knowledge of Language*. New York: Praeger.

Christophe, A., Guasti, M. T., Nespor, M., Dupoux, E. and van Ooyen, B. (1997). Reflections on prosodic boot strapping: its role for lexical and syntactic acquisition. *Language and Cognitive Processes*, **12**, 585–612.

Christophe, A. and Morton, J. (1998). Is Dutch native English? Linguistic analysis by 2-month-olds. *Developmental Science*, **1**(2), 215–19.

Christophe, A., Guasti, M. T., Nespor, M. and van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: the case of the head-complement parameter. *Developmental Science*, **6**, 213–22.

Colombo, J. and Bundy, R. S. (1983). Infant response to auditory familiarity and novelty. *Infant Behavior and Development*, **6**, 305–11.

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

Cutler, A., Mehler, J., Norris, D. and Segui, J. (1983). A language specific comprehension strategy. *Nature*, **304**, 159–60.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, **11**, 51–62. Le coup de grâce porté à la théorie de l'isochronie des intervalles inter-stress, et la mise en évidence d'une base phonétique et phonologique pour la sensation de rythme syllabique ou à stress.

Dehaene-Lambertz, G. and Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, **370**, 292–5.

Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, **384**(14 November), 159–61.

Fodor, J. (1975). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Geschwind, N. (1970). The organization of language and the brain. *Science*, **170**, 940–4.

Hebb, D. O. (1949). *Organization of Behavior*. New York: Wiley.

Hirsh-Pasek, K. and Golinkoff, R. M. (1996). *The Origins of Grammar: Evidence From Early Language Comprehension*. Cambridge, MA: MIT Press.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language Recognition*. Cambridge, MA: MIT Press.

Jusczyk, P. W., Rosner, B. S., Cutting, J. E., Foard, F. and Smith, L. B. (1977). Categorical perception of non-speech sounds by two-month old infants. *Perception and Psychophysics*, **21**, 50–4.

Kanwisher, N., McDermott, J. and Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, **17**(11), 4302–11.

Kluender, K. L., Lotto, A. J., Holt, L. L. and Bloedel, S. L. (1998). Role of experience for language-specific functional mapping of vowel sounds. *The Journal of the Acoustical Society of America*, **104**(6), 3568–82.

Kuhl, P. (1987). The special-mechanisms debate in speech research: categorization tests on animal and infants. In S. Harnad, ed., *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press, 355–86.

Kuhl, P. K. and Miller, J. D. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, **190**, 69–72.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. and Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, **255**, 606–8.

Ladefoged, P. (1975). *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.

Lenneberg, E. (1967). *Biological Foundation of Language*. New York, NY: Wiley.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech revised. *Cognition*, **21**, 1–36.

Liberman, M. and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, **8**, 240–336.

MacWhinney, B. (1987). *Mechanisms of Language Acquisition*. Hillsdale, NJ: Erlbaum.

Manrique, A. M. B. D. and Signorini, A. (1983). Segmental durations and rhythm in Spanish. *Journal of Phonetics*, **11**, 117–28.

Mazuka, R. (1996). How can a grammatical parameter be set before the first word? In J. L. Morgan and K. Demuth, eds., *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 313–30.

Mehler, J. (1981). The role of syllables in speech processing: infant & adult data. *Philosophical Transactions of the Royal Society*, **B295**, 333–52.

Mehler, J. and Dupoux, E. (1994). *What Infants Know*. Cambridge, MA: Blackwell.

Mehler, J. and Nespor, M. (2004). Linguistic rhythm and the development of language. In A. Belletti (ed.) *Structures and Beyond. The Carthography of Syntactic Structures*. Vol. 3. Oxford: Oxford University Press, pp. 213–22.

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

Meisel, J. M., (ed.) (1992). *The Acquisition of Verb Placement. Functional Categories and V2 Phenomena in Language Acquisition*. Dordrecht: Kluwer Academic Press.

Miller, G. A. (1951). *Language and Communication*. New York: McGraw-Hill Book Company Inc.

Molfese, D. L. and Molfese, V. J. (1979). Hemisphere and stimulus differences as reflected in the cortical response of newborn infants to speech stimuli. *Developmental Psychology*, **15**, 501–11.

Moon, C., Cooper, R. P. and Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, **16**, 495–500.

Morgan, J. L. and Demuth, K. (1996a). Signal to Syntax: an overview. In J. L. Morgan and K. Demuth, eds., *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 1–22.

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

Morgan, J. L., Meier, R. P. and Newport, E. L. (1987). Structural packaging in the input to language learning: contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, **19**, 498–550.

Morgan, J. L. and Saffran, J. R. (1995). Emerging Integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, **66**, 911–36.

Nazzi, T., Bertoncini, J. and Mehler, J. (1998). Language discrimination by newborns: towards an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, **24**(3), 756–66.

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

Nespor, M., Guasti, M. T. and Christophe, A. (1996). Selecting word order: the Rhythmic Activation Principle. In U. Kleinhenz, ed., *Interfaces in Phonology*. Berlin: Akademie Verlag, pp. 1–26.

Nespor, M., Mehler, J., and Peña, M. (2003). On the different role of vowels and consonants in language processing and language acquisition. *Lingue e Linguaggio*, 221–47.

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

Newport, E. L. and Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, **48**, 127–62.

Peña, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F. and Mehler, J. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences USA*, **10**, 11702–5.

Pike, K. L. (1945). *The Intonation of American English*. Ann Arbor, Michigan: University of Michigan Press.

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

Port, R. F., Dalby, J. and O'Dell, M. (1987). Evidence for mora-timing in Japanese. *Journal of the Acoustical Society of America*, **81**(5), 1574–85.

Premack, D. (1971). Language in chimpanzee. *Science*, **172**, 808–22.

Premack, D. (1986). *Gavagai!* Cambridge, MA: MIT Press.

Querleu, D., Renard, X., Versyp, F., Paris-Delrue, L. and Creprin, G. (1988). Fetal hearing. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, **29**, 191–212.

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

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

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

Segalowitz, S. J. and Chapman, J. S. (1980). Cerebral asymmetry for speech in neonates: a behavioral measure. *Brain and Language*, **9**, 281–8.

Seidenberg, M. S. and MacDonald, M. C. (1999). A probabilistic constraint approach to language acquisition and processing. *Cognitive Science*, **23**(4), 569–88.

Selkirk, E. O. (1984). *Phonology and Syntax: The Relation Between Sound and Structure*. Cambridge, MA: MIT Press.

Stevens, K. (2000). *Acoustic Phonetics*. Cambridge, MA: MIT Press.

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

Villringer, A. and Chance, B. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neuroscience*, **20**(10), 435–42.

Wanner, E. and Gleitman, L. R. (1982). *Language Acquisition: The State of the Art*. Cambridge, UK: Cambridge University Press.

Werker, J. F. and Tees, R. C. (1983). Developmental changes across childwood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, **37**, 278–86.

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

Wexler, K. and Culicover, P. (1980). *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.