Why is language unique to humans?

Jacques Mehler*†[1], Marina Nespor‡, Mohinish Shukla* and Marcela Peña*

* *International School for Advanced Studies, Trieste, Italy,* † *Ecole des Hautes Etudes en Sciences Sociales, Paris, France and* ‡ *Universita di Ferrara, Ferrara, Italy*

Abstract. Cognitive neuroscience has focused on language acquisition as one of the main domains to test the respective roles of statistical vs. rule-like computation. Recent studies have uncovered that the brain of human neonates displays a typical signature in response to speech sounds even a few hours after birth. This suggests that neuroscience and linguistics converge on the view that, to a large extent, language acquisition arises due to our genetic endowment. Our research has also shown how statistical dependencies and the ability to draw structural generalizations are basic processes that interact intimately. First, we explore how the rhythmic properties of language bias word segmentation. Second, we demonstrate that natural speech categories play specific roles during language acquisition: some categories are optimally suited to compute statistical dependencies while other categories are optimally suited for the extraction of structural generalizations.

2005 Percept, decision, action: bridging the gaps. Wiley, Chichester (Novartis Foundation Symposium 270) p 251–284

Linguists and psychologists have studied language acquisition; the former have elaborated the most sophisticated formal theories to account for how this unique competence arises specifically in humans. For instance, Chomsky (1980) formulated the Principles and Parameters theory (hereafter, P&P) to account for the acquisition of language given the poverty of the linguistic data the learner receives. In fact, infants acquire the grammatical properties of their language of exposure on the basis of partial and unreliable information. Babies, like adults, are confronted with incomplete or erroneous sentences.

P&P assumes that infants are born with 'knowledge' of Universal Grammar. This endowment includes genetically specified universal principles, that is, the properties shared by all natural languages. Moreover, the endowment specifies a number of binary parameters that capture those grammatical properties that vary systematically between groups of natural languages. For instance, there are groups of lan-

[1] This paper was presented at the Symposium by Jacques Mehler, to whom correspondence should be addressed.

guages that put Heads to the left of Complements while other languages put Complements to the left of Heads. The P&P theory attempts to identify such structural properties that are basic to natural language distinctions. Parameters can be thought of as switches that must be set to one of two possible positions to specify the properties of the language being learned. The linguistic input determines the particular value of a parameter.[2]

P&P has many virtues. First, by exploring the way in which natural languages are sorted into groups that share coherence for syntactic properties, P&P is one of the most productive theories ever developed within the linguistic domain, (see Baker [2001] for an accessible and fascinating account of the P&P proposal). Next, P&P also addresses the problem of language acquisition without making the simplifications common to alternative theories. For example, optimists claim that imitation is the privileged mechanism responsible for the emergence of grammatical competence. The P&P perspective is appealing because it is biologically realistic assuming that infants are equipped with a species-specific mechanism to acquire natural language that can be explored with the tools available to formal linguistics and to the explorations of cognitive neuroscience.

While P&P is certainly playing an important role in the domain of language acquisition, there is a second influential position that asserts that the surface properties of stimuli can bias the learner towards postulating syntactic properties for the incoming utterances. While the P&P theory was formulated with the precision necessary to allow us to evaluate it, the general learning device proposal appears to be somewhat less precise. Criticisms of proposals according to which general learning mechanisms are sufficient to explain language acquisition have been given by many theoreticians, see Chomsky (1959), Fodor (1975), Lenneberg (1967) and Pinker (1994) among many others. We will come back to this point below.

Recently, some attempts were made to show that speech signals contain hitherto ignored information to allow general learning accounts to explain how language is acquired. Mostly, these attempts are minor modifications of association, a mechanism that humans and animals share. Within this stream of research, the brain is

[2]To illustrate this, consider a child who hears mostly sentences with a Verb–Object order. The child, supposedly, obtains information automatically from the linguistic input to set the relevant word order parameter. If this were so, it would constitute a great asset, since fixing this word order parameter may 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 e.g. Italian ('piove', 'mangiano arance' etc.), or to a language whose sentences require overt mention of subjects e.g. English ('it is raining', 'they eat oranges'), supposedly gets information from the linguistic input to set the relevant pro-drop parameter.

regarded as a huge network that works in a Hebbian fashion (see Hebb 1949).[3] This may explain why many psychologists and neuroscientists have adopted a viewpoint that ignores the complexity of syntax and assumes that by focusing exclusively on speech perception and production, a functional account of how language is processed will follow. Undeniably, behavioural scientists have made great strides in the study of 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) arises in the infants' mind. Of course it is easier to study speech perception in babies or animals than trying to figure out how the human brain computes syntax, semantics and pragmatic's of utterances, something that animal brains cannot do. Psychologists who adopt a general learning device framework often assume that the mystery of syntax acquisition will disappear once we understand how infants just learn to extract the distributional properties of language (see Seidenberg & MacDonald 1999 among many others).

At this point we would like to point out that although there is a huge contrast between the two stances presented above there are many points of agreement as well. Investigators working in both the P&P tradition and in the general learning framework agree that some parts of grammar must be learned. Indeed, no one is born knowing Chinese, Malay or any other natural language. Each learner has to acquire the language spoken in his or her surrounds. What distinguishes the different positions is the scope and nature of learning they are willing to posit. P&P assumes an initial state characterized by 'knowledge' specific to a putative language module, i.e. Universal Grammar (UG). In contrast, general learning theoreticians assume that the initial state is characterized by learning principles that apply across the different domains in which organisms acquire knowledge. The general learning device is undeniably a powerful account to explain how organisms will use the surrounds to acquire behaviours that satisfy the organism's needs. Thus, it is an empirical issue whether just one of these theories is sufficient to explain how the human mind acquires its capacities including natural language.

Do we have accounts of how syntax can be acquired after the child has learned the lexicon of the language of exposure applying general learning mechanisms? Is there any evidence that the acquisition of syntax does not start until at least part of the lexicon is acquired? Depending on the answers we give to these questions, either the P&P model or the general learning model should be abandoned for syntax acquisition.

[3] Hebbian networks are a set of formal neurons synaptically connected but with connectivity values that change with functioning. If two neurons are active at the same time, the value of their connection increases. Otherwise, the value of the connection stays identical to what it was or decays.

So far, we have tried to highlight the positive aspects of both P&P and general learning mechanisms. However, problems arise with both frameworks. While the first tries to cope with language acquisition in a realistic sense the second focuses on the acquisition of, at best, toy-languages. P&P is problematic because of the many implicit assumptions that investigators make when trying to explain the acquisition of grammar. P&P was formulated with syntax acquisition in mind and researchers generally take for granted that infants, in one way or another, have already acquired the lexicon, before setting syntactic parameters. Presupposing that infants begin processing speech signals only when they start learning the lexicon justifies neglecting the study of language acquisition during the first year of life and explains why P&P investigators have mostly reported data from language production studies.

Data from animal experiments suggests that the vertebrate auditory system is optimally suited to process some of the linguistically relevant cues that speech affords. Thus, at least some properties of language could be acquired precociously from speech signals. Indeed, animals with auditory systems similar to our own tend to respond to speech patterns much like infants younger than eight months (see Kuhl 1987 and Ramus et al 2000 among many others). Apes, but also dogs, have 'lexicons' that can attain a few dozen words (Premack 1971, 1986). However, such abilities are insufficient to enable non-human animals to construct a grammar comparable to that of humans. Nonetheless, together with other pieces of evidence that we lay out below, we assume that the sensory capacity of many vertebrates licenses the processing of speech from the first year of life and, consequently, we should not neglect the aquisition that humans make during their first year. We show below that language acquisition begins with the onset of life. Indeed, several investigators, regardless of the position they defend, have found empirical evidence suggesting that the sound pattern of language are identified by very young infants and that some properties can be attested even in neonates. The sound pattern of speech contains cues that might bias language acquisition at different stages. As is becoming obvious, the viewpoints we presented above are complementary. Indeed, while rationalists and empiricists acknowledge the role of learning in language acquisition, the nature of learning conceived by each of the viewpoints is radically different. In the pages below we will try to show that it is desirable to keep in mind that only human infants use the acoustic properties of speech to acquire grammar. In order to explain how such uniqueness comes about, the theory that will eventually be preferred will be the one that fits best with biological processes.

We know that the uniqueness of syntax must be explored formally and explained with models that are biologically realistic. Indeed, we are confronting a human aptitude that will bloom under several types of impoverished learning environments. The linguistic input comes usually in the form of speech signals or, less often, in

the form of hand gestures as produced by deaf humans. Whether the learner is hearing and seeing, deaf or even blind, s/he will attain a grammar that is as rich and complex as we expect it in humans without sensory filters, see Klima & Bellugi (1979) and Landau & Gleitman (1985) amongst others.

Thus, not only do we have to account for the uniqueness of the human language ability but we also have to account for how language arises despite all the described impoverished conditions. The best way to attain such an aim is to use the specifications given in P&P to explain what needs to be learned and what may be mastered throught general learning procedures.

Chomsky (1980, 1986) and others have argued that conceiving acquisition of language from a P&P perspective will bring clarity to the field. However, the mechanisms for the setting of parameters in the P&P theory were seriously underspecified so as to make it hard to judge. In fact, Mazuka (1996) argues that, in its usual formulation, P&P contains a paradox (see below). Morgan et al (1987), Cutler (1994) and Nespor et al (1996) among others, have proposed some putative solutions to some of the problems arising within the P&P proposal. However, few proposals have explored how the infant evaluates and computes the triggering signals. Some recent results suggest that two-month-olds are sensitive to the prosodic correlates of the different values of the head-complement parameter (Christophe et al 1997, 2003).

In the early 1980s, some scholars like Wanner & Gleitman (1982) already foresaw some of the difficulties in the existing theories of grammar acquisition and proposed that *phonological bootstrapping* might help the infant out of this quandary. They held that some properties of the phonological system learnt by the child may help him/her to uncover lexical and syntactic properties. Some years later, Morgan & Demuth (1996) specifically added that prosody contains signals that can act as triggers and thus help the child learn syntax. Indeed, these authors conclude, as we do above, that the study of speech signals that can act as triggers is essential if we are to understand the first steps into language.

To overcome the poverty of the stimulus argument, innate dispositions were postulated. However, as pointed out above, the proposal for language acquisition is not sufficiently specific. Indeed, if an important part of the infant's endowment comes as binary parameters, we still need to understand how these are set to values that are adequate to the surrounding language. The general assumption was that by understanding a few words or simple sentences like '*drink the juice*' or '*eat the soup*' the child would generalize that in her/his 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 the mother always says *eat* in front of different foods, the child may understand that what she means is simply *food*! If the signals were to inform the child about lexical categories or word order, one could find a way out of this paradox. Before we know if this is a

valid 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 (e.g. Selkirk 1984, Nespor & Vogel 1986). These authors found systematic associations between these two grammatical levels, making plausible the notion that signals may cue the learner to postulate syntactic properties in an automatic, encapsulated fashion.

What is the infant learning during the first 18 months? Possibly, the answer is related to the infants' ability to perceive and categorize the cues that can act as triggers. Since these are supposed to function in an automatic and encapsulated way, supporters of the prosodic bootstrapping hypothesis are committed to the view that infants have 'learned' many aspects of their language before they begin to produce speech. These researchers have to give an account of the specific processes that occur during the first months of life. As we and others have argued, a parameter cannot be set after listening to a single utterance. Rather, properties of utterances are stored and the information is presumably used to set a parameter when it has become 'reliable'. Since some parameters can only be set after other grammatical properties have already been acquired (each of them requiring considerable information storage), we could perhaps understand the 'slow' pace of learning.

The sound of words is arbitrary as is clear from its variation attested across languages. On top of having to learn the identity of words, the child has to discover when a multi-syllabic utterance contains one or more words. Since most words are heard in connected speech, the infant has to rely on procedures to parse speech signals into its constituent morphemes. How does the infant parse speech to identify potential words? A proposal made by Saffran et al (1996) is that this can be achieved even by eight month olds, by computing the statistical properties of incoming speech signals. Thus, although we assume that UG is part of the infant's endowment and that it guides language acquisition, we also acknowledge that the statistical properties of the language spoken in the surrounds inform and guide learning. This is in contrast with the position of some theorists who argue that it is possible to explain even how grammar is acquired, exclusively on the basis of infants' sensitivity to the statistical properties of signals. How would such models stand up against real settings in which infants learn language from signals they receive? This issue was addressed by Yang (2005) who concluded that probabilities alone would not allow infants to converge to the words of the language given by the input.

The above presentation makes it clear that more data and research is needed to understand how the human biological endowment interacts with learning abilities during the first months of life. We are in a rather good position to do this because

during the last few years, new and fascinating results have been secured, allowing us to start having a more coherent picture of language acquisition. We will first explore whether the brain of newborn infants is specialized for language processing or whether this specialization arises as a consequence of language acquisition.

Innate dispositions for language?

Infants experience speech in noisy environments both before and after birth. Paediatricians tend to conjecture, as do naïve observers, that the racket infants experience after birth does not interfere with the processing of speech since they learn to focus on speech during gestation. But the womb is far from being the soundproof chamber that one might imagine; the womb is a very noisy place. Experiments with pregnant non-human vertebrates and volunteer pregnant women reveal that intra-uterine noise is as great as, if not greater than, the noise infants encounter after birth. The bowels, blood circulation, and all kinds of movements generate considerable noise (Querleu et al 1988). Thus, the womb is not the place to learn how to segregate speech from background noise. How then does the infant identify the signals that carry linguistic information? Why are music, telephone rings, animal sounds, traffic noises and other noises segregated during language acquisition?

Among the first researchers to focus on this issue we must mention Colombo & Bundy (1983) who found that young infants respond preferentially to speech streams as compared to other noises. This result, however, is difficult to appraise. There are zillions of noises out there and it is quite likely that infants might prefer some of them to the speech used by Colombo & Bundy (1983). We can always imagine that some melody is more attractive than a speech stream. Unfortunately few experiments have convincingly investigated this area. In an indirect effort, Mehler et al (1988) found that neonates and two-month olds process better, or more attentively, normal speech utterances as compared to utterances played backwards. The authors interpret their finding as showing that the infants brain preferentially processes speech, rather than non-speech stimuli. The uniqueness of this experiment resides in the numerous physical properties that these stimuli share, i.e. pitch, intensity and duration. However, in order to argue 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 or sneezes) more studies would be desirable.

Mehler et al's (1988) study pitted stimuli that could have been produced by the human vocal tract to stimuli that the human tract is incapable of producing. Thus, we ignore whether the infant's behaviour is determined by a speech vs. non-speech contrast or by a contrast between vocal vs. non-vocal-like sounds. Belin et al (2000)

have recently claimed that the human brain has an area that is devoted to processing conspecific vocal productions. Adults in an fMRI experiment listened to various speech and non-speech sounds (laughs, coughs, sighs, etc.) generated by the human vocal tract. The authors reported that sounds produced by the vocal tract elicit greater activation than non-vocal sounds bilaterally in non-primary auditory cortex. However, vocal sounds elicit greater activation than non-vocal sounds bilaterally along the superior temporal sulcus (STS). On the basis of these results they argue that there is a 'voice-region' much as there is a 'face-region'. Conceivably, under different experimental conditions, one could find areas that are selectively activated by speech-like stimuli that the human vocal tract could generate, as compared to similar stimuli that it could not generate. Does the brain have a localizer for processing human voices much as Kanwisher et al (1997) have proposed that faces are processed in the FFA (fusiform face area)? According to Belin and colleagues, human voice is processed in the STS. This conclusion may be premature and more experiments would be needed to be convincing.[4]

In our laboratory we have studied the specificity of the cortical areas devoted to processing different information-types, before any learning has occurred. Establishing that a brain area is a localizers of a function does not tell us how the area acquired this function. Our own efforts centre on the initial state of the cognitive system and of the brain structures that support it. Adults have already learned how to process and encode faces or human vocal tract productions, and might as a result have cortical tissue dedicated to this competence. Therefore, in order to distinguish aptitudes that arise as part of our endowment from those that arise as a consequence of learning, it is useful to investigate very young infants and, whenever possible, neonates. Indeed, during the first months of life, infants acquire many language specific properties (see Werker & Tees 1984, Kuhl et al 1992, Mehler & Dupoux 1994, Jusczyk 1997).

Standard neurological teaching tells us that the left hemisphere (LH) is more involved in language representation and processing than the right hemisphere (RH) (see Dronkers 1996, Geschwind 1970, Bryden 1982, among many others, but see

[4] To establish that the FFA is an area that is specifically responsive to faces, Kanwisher had to test many other stimuli and conditions. Gauthier et al (2000) have challenged the existence of the FFA showing that this area is also activated by other sets of stimuli whose members belong to categorized ensembles even though they are not faces. Indeed, Gauthier and her colleagues showed that when Ss learn a new set before the experiments, its members then activate the FFA. Gauthier argued that her studies show that the FFA is not a structure uniquely 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 the FFA much better than Belin's voice area 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.

also Gandour et al 2002). Are infants born with specific LH areas devoted to speech processing or is LH specialization solely the result of experience? The response to this question is still tentative. Some studies report that infants are born with speech processing abilities similar to those of experienced adults. For instance, infants discriminate all the phonetic contrasts that arise in natural languages (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 processing capacities explain these early abilities that humans share with other organisms (Jusczyk 1997, Jusczyk et al 1977, Kuhl & Miller 1975). Thus, though it is conceivable that humans are endowed with a species-specific disposition to acquire natural language, we lack the data that might answer whether we are born with cortical structures specifically dedicated to the processing of speech.

Experimental psychologists devoted substantial efforts to establish whether LH superiority is the consequence of language acquisition or whether language is mastered because of this cortical specialization. Most studies have found an asymmetry in very young humans (Best et al 1982, Bertoncini et al 1989, Segalowitz & Chapman 1980). A few ERP studies have also found trends for LH superiority in young infants (Molfese & Molfese 1979, Dehaene-Lambertz & Dehaene 1994). Both the behavioural and the ERP data suggest that LH superiority exists in the infants' brain.

Below we review results obtained using more advanced imaging methods to study functional brain organization in newborn infants. Several methods are being pursued in parallel. A few groups have begun to study healthy infants using fMRI (Dehaene-Lambertz et al 2002). In the following section, we focus on recent results we obtained with Optical Topography (OT).

Brain specialization in newborns: evidence form OT

Optical Topography is a method derived from the Near Infrared Spectroscopy (NIRS) 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.[5] In particular, it allows one

[5] This non-invasive device uses near-infrared light to evaluate how many photons are absorbed in a part of the brain cortex following stimulation. Like fMRI, it estimates the vascular response in a given area of the cortex. As fMRI, it estimates changes in deoxHb, however, it also gauges changes in oxyHb correlated with stimulation. Like fMRI, its time resolution is poorer than that of ERP. Our device uses bundles of source and detector fiber optics that are applied to the infants' head. The source fibers deliver near-infrared light at two wavelengths. One of the wavelengths is better absorbed by oxyHb, while the other is better absorbed by deoxyHb.

to estimate the concentration of oxy-haemoglobin (oxyHb), deoxy-haemoglobin (deoxyHB) and total haemoglobin over a given area of the brain.

Peña et al (2003) used a NIRS device to evaluate whether the neonate's brain is specifically tuned to human speech. Using sets of light emitting fibres and light detecting fibres, and two wavelengths, one can observe how the cortex responds to stimuli on homologous areas of the LH and RH. We placed the probes so as to measure activity over the RH and LH temporal and parietal areas. Participants were tested with Forward Speech (FW): infants heard sequences of 15 seconds of connected French utterances separated from one another by periods of silence of variable duration, i.e. from 25–35 s. In another condition, Backward Speech (BW), infants were tested as in the FW condition but with the speech sequences played backwards. In this second condition the speech signal was converted from FW to BW using a speech waveform editor. Ten such blocks of FW and BW conditions were used. Finally, in a control condition, infants were studied in total silence for a duration identical to that of the above conditions.

The results show that, as in adults, the haemodynamic response begins four to five seconds after the infant receives auditory stimulation. This time-locked response appears more clearly for the oxyHB than for the deoxyHB.[6] The results also show that roughly five seconds after the presentation of the FW utterances, a robust change in the concentration of totalHB takes place over the temporo-parietal region of the LH. Interestingly, the concentration of totalHb is relatively lower and comparable both in the BW and in the Silence conditions. Thus only forward speech gives rise to a significant increase in total Hb over the LH, while BW speech does not give rise to a significant increase in total HB in any of the channels. While the acoustic energy is identical in the FW and BW sentences, and their spectral properties are mirror images of each other, the brain responds very differently to the acoustic pattern that can be produced by the human vocal tract in contrast to that which cannot.

The reported results suggest that the brain of the newborn infant responds differently to natural and backward speech. To understand the singularity of this result, it may be useful to mention that in a pilot study we found that monolingual adults who were tested with materials similar to those used with 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. FW and BW speech differ in terms of their timing patterns. Indeed, final lengthening appears to be a universal property of natural language. Thus, only BW utterances

[6] We now know that a better choice of wavelengths would have permitted us to avoid this problem.

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 backwards. The vocal tract cannot produce backward speech. Since infants cannot produce forward speech either, they might have ignored the contrast between the BW and FW conditions (Liberman & Mattingly 1985). However, since the neonate's brain responds differently to FW and BW we infer that, in some sense, the infants' brain has become attuned to the difference between natural and unnatural utterances. 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. We are replicating this result using an improved NIRS device. The new machine has wavelengths that are better suited to track vascular responses and moreover it is equipped with probes that are designed to fit better on the infant head.

It is not only necessary to replicate the above result but it is also necessary to have a better theoretical grasp of what these results entail. If the LH dominance already observed in neonates is viewed as an emergent evolutionary module then we ought to explore whether asymmetrical patterns of activation to FW and BW speech are also found in non-human primates or even more primitive vertebrates. As a matter of fact, work with monkeys exists and suggests that their behaviour is similar to that of infants, when exposed to FW and BW speech. In a series of studies comparing the human newborn and the adult cotton-top tamarin monkey for their behavioural responses to FW and BW speech, Ramus et al (2000) showed that, like infants, tamarins discriminate two different languages (Japanese and Dutch) when the utterances are played forwards but fail to do so when the utterances are played backwards. The ability to behave like infants in the FW and BW conditions is remarkable since tamarins will never develop speech. This outcome ought to temper the desire to conclude that the infant results are based on a species-specific system to process natural speech. Indeed, the observed specialization may have arisen much before language arose. Many vertebrates produce cries and vocal noises in this way and a specialized module might have evolved to discriminate such sounds from other sounds that cannot be generated by these kinds of vocal tracts.

Obviously, the advent of imaging studies using neonates will permit 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 the vocal tract of any vertebrate species. In the meantime, these studies have shed some light on complex issues that were hard to study with more traditional behavioural methods. To close this section let us just remind the reader of a study carried

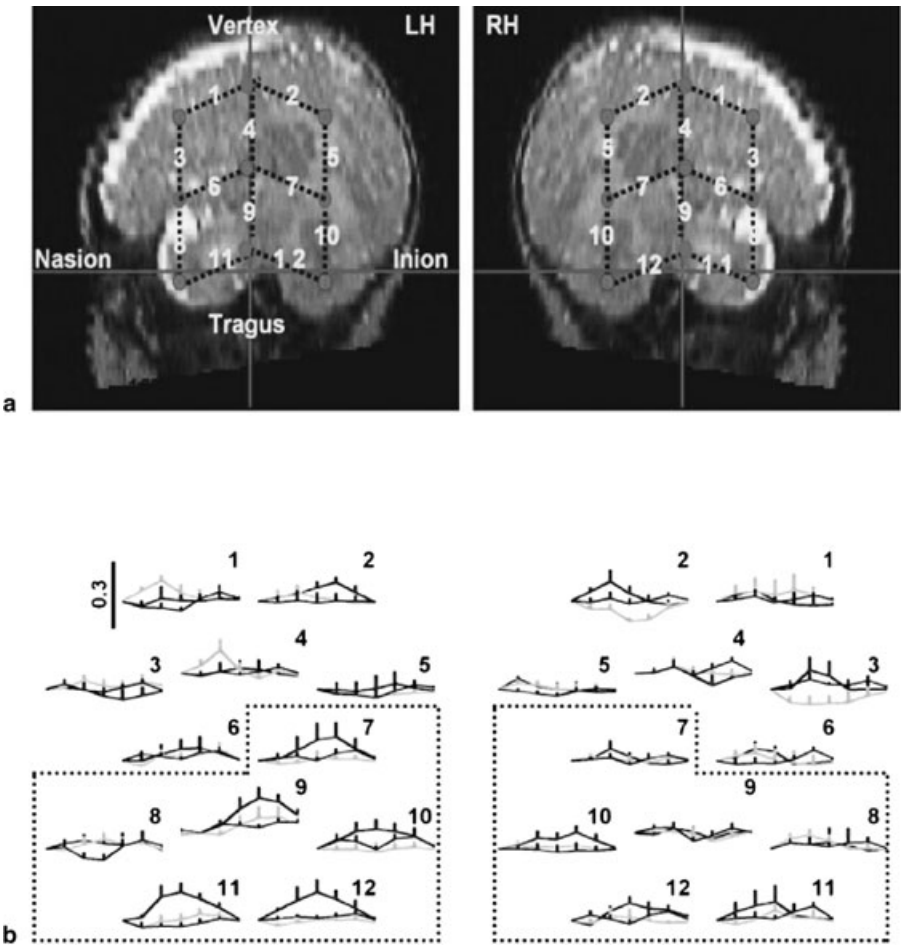

FIG. 1. Changes in total haemoglobin for newborn Italian infants. Each infant contributes more than three blocks in each one of the conditions. All blocks are summed across infants. (a) Indicates how the probes were placed over the left and right hemispheres (LH, RH). (b) Results showing the activity recorded over each one of the hemispheres. Darkest grey, forward speech; lightest grey, backward speech; intermediate grey, silence. Reproduced by permission of Peña et al (2003).

out with three-month-olds tested with FW, BW and Silence using an fMRI device. Dehaene-Lambertz et al (2002) showed that cortical regions were active well before the infant has acquired the native language.

Let us now turn to another property that is essential to understand the first adaptations that the human infant makes to speech stimuli, that is rhythm. Not only did

linguists use the notion of rhythm to sort languages into groups or classes but, independently, developmental psychologists discovered that rhythm is the very first adjustment that infants make to the maternal language.

Neonates use rhythm to tune into language

The notion of rhythm relates to the relative duration of constituents in a sequence. What, we can ask, are the elements responsible for the characteristic perception of the rhythm of a language? Three constituents, or atoms, of different size have been proposed to be roughly isochronous in different languages, thus giving rise to rhythm: syllables, feet and morae (Pike 1945, Abercrombie 1967, Ladefoged 1975). Syllables have independently been construed as a basic constituent or atom for both speech production and speech comprehension (Levelt 1989, Mehler 1981, Cutler & Mehler 1983). Infants begin to produce syllables several months after birth, with the onset of babbling. However, the infant may rely on syllables to process speech before s/he produces syllables. If so, it should be possible to find indications that neonates process syllables in linguistic-like ways.[7] Bertoncini & Mehler (1981) explored this issue using the non-nutritive sucking technique and showed that very young infants distinguish a pair of syllables that differ only in the serial order of their constituent segments, e.g. *PAT* and *TAP*. The infants, however, failed to distinguish a pair of items, i.e. *TSP* and *PST*, which were derived from the previous ones by replacing the vowel [a] by the consonant [s]. This editing of the 'good' syllables transforms the new items into 'marked' or 'bad' syllables. To understand the infants' failure to distinguish this pair, we ran a control experiment, in which infants were presented with the same 'marked' syllables inserted in a context of vowels, i.e. *UPSTU* and *UTSPU*, that generated two bi-syllabic, well-formed speech sounds. When these sequences were presented to the infants, discrimination ability was restored. This experiment suggests that the infant discriminates items presented in a linguistic-like context but s/he neglects those constructs in other acoustic contexts.

[7] 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 positions 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, in particular [r] and [l]. Thus, a syllable may not contain more than one vowel (or a dipthong). CV is the optimal syllable, i.e. the onset is 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, especially if none of the consonants is sonorant.

As we mentioned in footnote 5, some languages, (e.g. Croatian and some varieties of Berber) allow specific consonants to occupy the nuclear position of the syllable. For instance, in Croatian, *Trieste*, the Italian city, is called *Trst* where [r] is the nucleus. This is not an exceptional case in the language. Indeed, the word for 'finger' is *prst*, the word for 'pitcher' is *vrča*, and 'square' or 'piazza' is *trg*. Why then did the infants respond as they did in the results reported in Bertoncini & Mehler (1981)? Why did the infants fail to treat *PST* and *TSP* as different syllables? One explanation may be that we tested rather old (i.e. two-month-olds) infants who had already gained considerable experience about the surrounding language. All the infants had been tested in a French environment; it is possible that the stimuli were already considered inappropriate for their language and thus they neglected the ill-formed stimuli. An alternative explanation that still needs to be explored is that PST and TSP are non-standard syllables in any language, including the ones named above. To the best of our knowledge there is no language that allows [s] as a syllabic nucleus. We predict that infants have no difficulty 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 more than 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 ensure 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.

Bijeljac-Babic et al (1993) had already claimed that very young French raised infants attend to speech using syllabic units; that is, units that are related to the rhythmical pattern instantiated in that language. These authors showed that infants distinguish lists of bisyllabic items from a list of trisyllabic 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 either the number of syllables or just the number of vowels is what counts. We focused on syllables rather than on feet or morae because of the total absence of studies that explored whether neonates can also represent those units. Below we will 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 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 (Moon et al 1993, Mehler et al 1988). 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 in the utterances of a language that is occupied by vowels (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 (Nazzi et al 1998).

The variability of the intervocalic interval (i.e. ΔC, the standard deviation of the intervocalic intervals) also plays an important role in explaining the infants' behaviour. 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 the three possible units maintains isochrony in the speech stream: stress-timed rhythm, syllable-timed rhythm and mora-timed rhythm (Pike 1945, Abercrombie 1967, Ladefoged 1975). However, once exact measures were carried out, isochronous units were not found (see Dauer 1983, Manrique & Signorini 1983, but see Port et al 1987 who have claimed that there is a fair amount of isochrony for morae). 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 (1999)'s definition of rhythm on the basis of ΔC and %V divides languages exactly into those three intuitive classes, as shown in Figure 2.

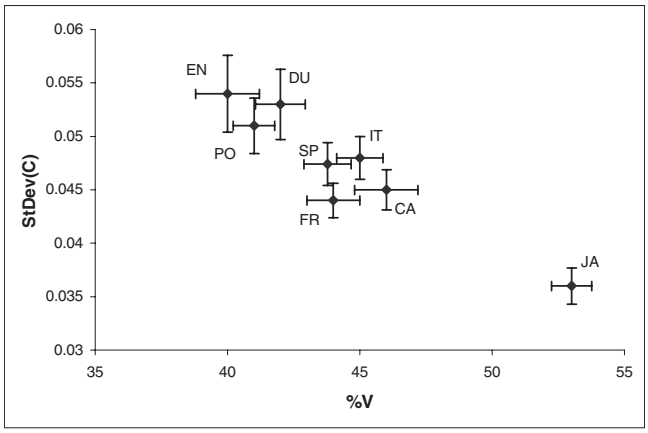

FIG. 2. %V is the mean proportion of the utterances in a language that is occupied by vowels and ΔC or StDev(C) is the standard deviation of the consonantal intervals. The plot incorporates eight languages spoken by four female speakers. Each speaker utters 20 sentences (each language is represented by 20 utterances). The distribution of the languages is compatible with the notion that they can be grouped into three classes as predicted by linguists' intuitions. Reprinted from Ramus et al (1999), with permission from Elsevier.

A language with a high %V and a low 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*).[8] Romance languages, as depicted in Figure 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 (e.g. *prêt*, *prato*) and, at least in some Romance languages, even codas contain more than one consonant (e.g. *tact*, *parc)*. However, fewer syllable types are allowed in Romance languages than in stress-timed languages like Dutch or English. Indeed, while in Romance languages the typical syllabic repertoire ranges from six to eight syllables, Germanic languages have over sixteen 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, and thus many consonant clusters, tend to have a smaller proportion of the utterances taken up by vowels. Interestingly, one could conclude that if a larger number of languages were included in Figure 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 favour of a clustering. Indeed, the syllable repertoires come in groups. Up until now, we have considered languages that have two or three syllable types (Hawaiian, Japanese, etc.), six to ten syllable types (Spanish, Greek, Italian, etc.) and languages that have sixteen 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 borne out or not, and if so, how many classes there are.

The conjecture we make is that rhythm, as defined by Ramus et al (1999), is sufficient to explain all the behavioural results showing that languages cluster into a few classes. Indeed, Ramus (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 available empirical findings involving neonates. This outcome sustains our resolve to pursue this line of investigation. Indeed, it seems

[8] Or geminates as in the word *Sapporo*.

unlikely that linguistic rhythm would be so salient for the neonate without having any further impact on how language is learned.

The first known adjustment to the surrounding language the neonate makes concerns rhythm. The processing of linguistic rhythm changes over the first two months of life. Mehler et al (1988) remarked that while American two-month-olds fail to discriminate Russian from French, one-week-old French infants successfully discriminate not only Russian from French but also English from Italian suggesting that by two 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 & Morton (1998) further investigated this issue, testing two-month-old and four-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 involves one familiar and one novel type of rhythm. The second switch is not discriminated because neither language has a rhythm that is familiar to the infant. To buttress their interpretation, Christophe & Morton (1998) also tested the behaviour of the 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 languages foreign to 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.

We hope to complement the above behavioural research with brain-imaging methods. If we succeed, this will provide more information to decide whether learning and development of language require a passage through an attention-drawing device based on rhythm. But even before we obtain such data we have to raise the following central question: Why are infants interested in rhythm even before the elementary sound patterns of utterances attract their attention?[9] What information does linguistic rhythm provide to render it so relevant for language acquisition? We have implemented two procedures to answer these questions. First, we have tried to gather data using optical topography 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 might play a role in the setting of syntactic parameters, and also whether it might be exploited in segmentation, as described in the following sections, see also Mehler & Nespor (2004).

[9] Werker & 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.

Segmenting the speech stream

Ramus (1999) conjectured that language rhythm provides the infant with information about the richness of the syllabic repertoire of the language of exposure (cf. Dauer 1983 and Nespor 1991). For the sake of argument, we assume that the infant gains this type of information from rhythmic properties in the signal. What could the use of such information for the language-learning infant be? What benefit could a baby draw once s/he learns that the number of syllable types is four, six or 16? Will such information help perception of speech? Or will such information be essential in mastering the production routines or elementary speech acts? We cannot answer these questions in detail. However, there is 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 s/he has prior knowledge of the number of types or items in the set to be learned? We can give an indirect answer to this question by looking at lexical acquisition. Infants appear to 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 necessary given that infants acquire thousands of words without, to the best of our knowledge, requiring special signals about word types? However, there is a tentative 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 a bias to look for large or smaller lexical items in the language of exposure (Nespor & Mehler 2003). This simple procedure may prove important to understand how infants identify potential words since we know that most words come in packages of fluent, continuous streams.

It is too early to decide whether infants make use of the above bias. Many other conjectures have been proposed. For instance, it has been suggested that many words are first learned when the mother produces them in isolation. Others have suggested that infants focus on the onset and the end of all the utterances they receive. This might allow them to isolate recurrent items. None of these conjectures should be ruled out. However, there is a proposal that infants parse continuous streams by finding dips in transition probabilities between some syllables and high transition probabilities between other syllables. (Saffran et al 1996).

Can rhythm also help segmenting the continuous speech stream? Nespor et al (2003) 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 word-like constituents while infants who listen to a language whose %V is below 40% will tend to search for shorter units. This follows from the fact that the syllabic repertoire in, e.g. Japanese, is very limited, which entails that monosyllables

will be rare and long words will be very frequent. We assume that speakers are unwilling to put up with polysemy to an extent that would threaten communication. Supposedly, languages are designed to favour rather than to hinder communication. In fact, 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 syllable types. Hence, it is easy to understand why 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 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 useful to understand the underlying computational processes. Unfortunately, at this time, we have no results that might 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 the speech stream, and/or whether we are born to be powerful statistical machines that allow for small differences in rhythm between classes of languages to be detected. Regardless of how the properties that characterize rhythmic classes are identified, our conjecture is that the trigger that leads 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. Regardless of the putative role of rhythm we acknowledge that statistical process are powerful and well attested as an instrument for segmentation while rhythm is only very indirectly related to segmentation: it is relevant for the infants and it predicts the mean length of words in languages according to their syllabic structure.

Saffran et al (1996) and Morgan & Saffran (1995) have revived the view that statistical information plays a central role in language acquisition. Indeed, Miller (1951) had already postulated that the statistical properties of language could help process signals and thus favour language acquisition. Connectionism has also highlighted the importance of statistics for language learning; it postulates that the language learner can be viewed as a powerful statistical machine. We acknowledge 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 nine-month-old infants confronted with unfamiliar monotonous artificial speech streams tend to infer word

boundaries through statistical regularities in the signal. A word boundary is postulated in positions where the transitional probability (TP)[10] drops between one syllable and the next.[11] Participants familiarized with a monotonous stream of artificial speech recognize trisyllabic items delimited by dips in TP. As an example, imagine that *puliko* and *meluti* are items with high TPs between the constituent syllables. If participants 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 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.[12]

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 intervocalic intervals. We also acknowledged the powerful role that statistics plays in helping determine early properties present in the speech stream. However, we also noticed that rhythm as defined in Ramus et al 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.

Rhythm, signals and triggers

Developmental psycholinguists and students of adult language perception and production have tried to evaluate whether the rhythmic class to which a language belongs is related to phonological units that are highlighted during processing, see Cutler (1993). More recently, linguists and psycholinguists have started exploring whether phonological properties related to syntax can guide the infant in the setting of the essential parameters necessary to acquire the grammar of the language. We

[10] Transition probability between two syllables is synonymous with the conditional probability that the second syllable will occur immediately after the first one.

[11] Saffran, Aslin & Newport (1996) use streams that consist of artificial CV syllables that are assembled without pauses between one CV and the next. All syllables have the same duration, loudness and pitch. TPs between adjacent syllables (within trisyllables) 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.

[12] One divergence between the results reported by the Rochester group and our own concerns the computation of TPs on the consonantal and vocalic tiers. Native English speakers can use both tiers to calculate TPs (Newport & Aslin 2004). Our own Ss, regardless of whether they are native French or native Italian speakers, can only use the consonantal tier. Notice that Newport & Aslin use only two families, generating repetitions that we did not allow (see main text below).

are presently exploring to what extent linguistic rhythm relates to the processes that leads to the discovery of the non-universal properties of his/her native syntax.

Our proposal is to integrate P&P 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 parameters has not been sufficiently explored. In fact, while signals may give a cue to the value of a certain parameter, general learning mechanisms may play a role in establishing the validity of such a cue. For instance, in order to decide whether complements in a language 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 are, by and large, of one type or another, i.e. they are either Head-Complement (HC) or Complement-Head (CH).[13] There are languages, however, in which the word order in a specific phrase can be different from the standard word order. Since the pre-lexical infant ignores whether exceptions of this kind weaken the information that overall prominence provides, there must be some mechanism for her/him to detect such cases. In all likelihood, statistical computations allow the infant to discover and validate the most frequently used phonological pattern that can act as a cue to the underlying syntax (Nespor et al 1996). Indeed, even an infant who is exposed to a regular language (as to the HC order) may occasionally hear irregular patterns, e.g. 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 on the adequate setting.

Let us focus more closely on the case of the HC parameter. In the great majority of languages, the setting of this parameter simultaneously specifies the relative order of heads and complements and thus of main clauses with respect to subordinate clauses. That children start the two-word stage without making mistakes in word order suggests that this parameter is set precociously (Bloom 1970, Meisel 1992). In addition, even prior to this, babies 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 such evidence a scenario in which the infant finds how to set basic parameters prior to, or at least independently of the segmentation of the speech stream into words seems sensible to explore. If the child sets parameters before learning the meaning of words,

[13] For example, in languages in which the verb precedes the object, subordinate clauses follow main clauses. In contrast, in languages in which the verb follows the object, subordinate clauses precede the main clause. Other ordinal properties also correlate with the HC or CH structure of languages; for a more technical definition of the notion of the head-complement parameter see, e.g. Haegeman (1994).

prosodic bootstrapping might 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 which word in a pair of words functions as the head and which as the complement, it already knows how they are ordered. If you know how they are ordered, the parameter becomes pointless for the purposes of acquisition. 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 provide a cue to the value of this parameter? Rhythm, in language as in music, is hierarchical in nature (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 nuclei) with respect to other syllables reflects some aspects of syntax. In particular, in the phonological phrase,[14] rightmost main prominence is characteristic of HC languages (such as English, Italian or Croatian) while leftmost main prominence characterizes CH languages (such as 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 (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 (2003) carried out a discrimination task using resynthesized 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 within the phonological phrase, an aspect that is crucial for us.[15] The experiment used delexicalized sentences pronounced by the same voice.[16] Six to 12-week old infants discriminated French from Turkish. It was concluded that infants discriminate the two languages only on the basis of the different location of the

[14] The phonological phrase is a constituent of the phonological hierarchy that includes the head of a phrase and all its function words, e.g. articles, prepositions and conjunctions; for a more technical definition see Nespor & Vogel (1986).

[15] The effect of the resynthesis is that all segmental differences are eliminated.

[16] Sentences were synthesized using Dutch diphones with the same voice.

main phonological phrase 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.

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 (1998), 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 allowing for an increase in the number of messages with different meanings that can be transmitted. Even if this explanation is correct, the reason languages necessarily include both vowels and consonants may have additional functional roles. Indeed, Nespor et al (2003) have proposed that vowels and consonants play a different functional role in language acquisition and language perception. Consonants are intimately linked to the lexicon structure, while vowels are 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 most frequent segmental system in the languages of the world has five vowels and around 20 consonants. But languages with just three vowels are also attested and historical linguists working on common ancestors of different languages have posited two or even one single vowel for proto-Indo-European. However, languages attested today have at least two vowels. For example, the Tshwizhyi and Abzhui dialects of Abkhaz contrasts only /a/ and /i̲/, with significant allophony.[17]

A widespread phenomenon in the languages of the world is vowel reduction in unstressed positions. Languages like English, in which unstressed vowels are centralized to schwa, thereby losing their distinctive power, represent an extreme case. Another widespread phenomenon is vowel harmony, whereby all the vowels in a certain domain share some features. No comparable phenomena affect 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

[17] Some linguists claim that it is possible to posit only *one* vowel in some Abkhaz dialects, though the general consensus seems to be that that is stretching things a bit.

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 tend to be perceived categorically, vowels do not (see Kuhl et al 1992, Werker & Tees 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 of 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 native language are related to vowels rather than to consonants. Indeed, several pieces of evidence can be advanced to buttress this assertion. Bertoncini et al (1988) showed that neonates 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.[18] However, two-month-olds show a response to both, i.e. 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. We can conclude that the first representation privileges vowels, but that by two months of age vowels and consonants are sufficiently well encoded as to yield a similar phonological representation. Similarly, while by six months of age infants respond preferentially to the vowels of their native language,[19] Werker & Tees (1984) have shown that convergence to native consonants happens later: consonantal contrasts that are not used in the native language are still discriminated before eight months and are neglected only a few months later. 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 mention below, vowels and consonants, even when they are equally informative from a statistical point of view, are not exploited in similar ways.

[18] 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 is neglected and so is the introduction of [da] after habituation with the second set.

[19] American infants respond preferentially to American vowels as compared to Swedish vowels while Swedish infants respond preferentially to Swedish vowels compared with English ones (Kuhl et al 1992).

Newport & Aslin (2004) used a stream of synthetic speech consisting of CV syllables of equal pitch and duration in which 'words' are characterized only by high TPs between the consonants, while vowels change in the different instantiations of a 'word'. The authors showed that participants have no difficulty segmenting such streams.[20] We replicated this robust finding with Italian and French-speaking participants (Peña 2002, Bonatti et al 2005). In a similar experiment in which the statistical dependences were carried by vowels while the intervening consonants vary, the participants tested at Rochester were able to segment the streams while our participants (French, Spanish or Italian native speakers for the different experiments) failed to segment the stream into constituent 'words'. There are several differences between the English language experiment and the ones run using Italian or French. First, the streams Newport and Aslin used to test the vowel tier and the consonantal tier were not comparable to the ones used in Bonatti et al (2005). As we pointed out before, they used only two families to carry out their experiments while we used three. This means that they were obliged to repeat families while we carefully avoided such repetitions. The repetition of families might allow a repetition detection mechanism to intervene, see Endress et al (2005). This, of course, would only show that repetition detection promotes segmentation and not that the statistical dependencies incorporated in the vowel tier are responsible for the behaviour. If we are right, participants can use statistics to segment streams on the basis of the consonantal tier but not on the basis of the vowel tier.

On the basis of our experiments we conclude that 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, we can ask why this should 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, i.e., *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 ensure 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. Identification of the rhythmic class, we argued, provides information about the syllable repertoires, i.e. a part of phonology. Moreover, it gives information

[20] Thus, if a word has the syllables C-, C′-, C″ with the consonants that predict the next one exactly, regardless of the vowels that appear between them, the word in question will be preferred to a part word like C″-,C*-, C** (where 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*.

about the mean length of words in the language (see Nespor et al 2004). In addition, the information carried by vowels relates to the location of the main prominence within the phonological phrase. As was argued above, this prominence is related to a basic syntactic parameter.

Conclusion

In this paper, we have argued that both innate linguistic structure and general learning mechanisms are essential to our understanding of the acquisition of natural language. Theoretical linguists have focused their attention on the universal principles or constraints that delimit the nature of our endowment for language. Psychologists have explored how the child acquires the language of exposure, without showing much concern for the biological underpinnings of this process. After scrutinizing the potential limitations of both positions, we have pleaded in favour of the integration of the two approaches to improve our understanding 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 highlight the importance of the teaching of formal linguistics and explore 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. This has also obliged us to explore how signals can drive rule-like computations, see Peña et al (2002). We also laid out a proposal of how rhythm can guide the learner towards the basic properties of the language's phonology and syntax. In addition, we have 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 highlight concerns the attested acoustic capacity of non-human vertebrates to discriminate and learn phonetic distinctions (see Kluender et al 1998, Ramus et al 2000). They also have the ability to extract and use the statistical properties of the stimulating sequences in order to analyse and parse them into constituents (M. Hauser, personal communication). These results suggest that humans and other higher vertebrates can process signals in much 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 non-universal 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 cues that the signals provide. How can this argument be sustained when we have just acknowledged that human and non-human vertebrates processes acoustic signals in a similar fashion? The reason is that 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 towards 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.

Acknowledgements

The research presented in this article is supported in the frame of the European Science Foundation EUROCORES programme *The Origin of Man, Language and Languages*, by the HFSP grant RGP 68/2002 and by the Regione Friuli—Venezia Giulia (L.R. 3/98). We are grateful to Judit Gervain for suggesting many and varied changes to our manuscript.

References

Abercrombie D 1967 Elements of general phonetics. Edinburgh University Press, Edinburgh

Baker M 2002 The atoms of language: The mind's hidden rules of grammar. Basic Books, New York

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B 2000 Voice-selective areas in human auditory cortex. Nature 403:309–312

Bertoncini J, Bijeljac-Babic R, Jusczyk PW, Kennedy LJ, Mehler J 1988 An investigation of young infants' perceptual representations of speech sounds. J Exp Psychol Gen 117:21–33

Bertoncini J, Mehler J 1981 Syllables as units in infant perception. Infant Behav Dev 4:247–260

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

Best CT, Hoffman H, Glanville BB 1982 Development of infant ear asymmetries for speech and music. Percept Psychophys 31:75–85

Bijeljac-Babic R, Bertoncini J, Mehler J 1993 How do four-day-old infants categorize multisyllabic utterances? Dev Psychol 29:11–721

Bloom L 1970 Language development: form and function in emerging grammars. MIT Press, Cambridge, MA

Bloom P 2000 How children learn the meaning of words. MIT Press, Cambridge, MA

Bonatti L, Peña M, Nespor M, Mehler J 2005 Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. Psychol Sci 16:451–459

Bryden MP 1982 Laterality: Functional asymmetry in the intact brain. Academic Press, New York

Chomsky N 1959 A review of B.F. Skinner's Verbal Behavior. Language 35:26–58

Chomsky N 1980 Rules and representations. Columbia University Press, New York

Chomsky N 1986 Knowledge of language: Its nature, origin and use. Praeger, New York
Christophe A, Guasti MT, Nespor M, van Ooyen B 2003 Prosodic structure and syntactic acquisition: The case of the head-complement parameter. Developmental Sci 6:213–222
Christophe A, Morton J 1998 Is Dutch native English? Linguistic analysis by 2-month-olds. Developmental Sci 1:215–219
Christophe A, Nespor M, Guasti MT, van Ooyen B 1997 Reflections on phonological bootstrapping: its role in lexical and syntactic acquisition. In: GTM Altmann (Ed.), Cognitive models of speech processing: a special issue of language and cognitive processes. Lawrence Erlbaum, Mahwah, NJ
Colombo J, Bundy RS 1983 Infant response to auditory familiarity and novelty. Infant Behav Dev 6:305–311
Cutler A 1994 Segmentation problems, rhythmic solutions. Lingua 92:81–104
Cutler A, Mehler J, Norris D, Segui J 1983 A language-specific comprehension strategy. Nature 304:159–160
Cutler A, Mehler J 1993 The periodicity bias. J Phonetics 21:103–108
Dauer RM 1983 Stress-timing and syllable-timing reanalysed. J Phonetics 11:51–62
Dehaene-Lambertz G, Dehaene S, Hertz-Pannier L 2002 Functional neuroimaging of speech perception in infants. Science 298:2013–2015
Dronkers NF 1996 A new brain region for coordinating speech articulation. Nature 384:159–161
Endress AD, Scholl BJ, Mehler J 2005 The role of salience in the extraction of algebraic rules. J Exp Psychol Gen 134:406–419
Fodor J 1975 Language of thought. Crowell, Scranton, PA
Gandour J, Wong D, Lowe M, Dzemidzic M, Satthamnuwong N, Tong Y et al 2002 A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. J Cogn Neurosci 14:1076–1087
Gauthier I, Skudlarski P, Gore JC, Anderson A W 2000 Expertise for cars and birds recruits brain areas involved in face recognition. Nat Neurosci 3:191–197
Geschwind N 1970 The organization of language and the brain. Science 170:940–944
Haegeman L 1994 Introduction to government and binding theory (Blackwell textbooks in linguistics, No 1). Blackwell Publishers, Oxford
Hebb DO 1949 The organization of behaviour. John Wiley, Chichester
Hirsh-Pasek KA, Golinkoff RM 1996 The origins of grammar: evidence from early language comprehension. MIT Press, Cambridge, MA
Jusczyk PW 1997 The discovery of spoken language. MIT Press, Cambridge, MA
Jusczyk PW, Rosner BS, Cutting JE, Foard CF, Smith LB 1977 Categorical perception of non-speech sounds by 2-month-old infants. Percept Psychophys 21:50–54
Kanwisher N, McDermott J, Chun MM 1997 The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311
Kuhl PK 1987 The special-mechanisms debate in speech research: categorization tests on animals and infants. In S Harnad (Ed), Categorical perception: The groundwork of cognition. Cambridge University Press, Cambridge, p 355–386
Kuhl PK, Miller JD 1975 Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. Science 190:69–72
Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B 1992 Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255:606–608
Klima E, Bellugi U 1979 The signs of language. Harvard University Press, Cambridge, MA
Kluender KR, Lotto AJ, Holt LL, Bloedel SL 1998 Role of experience for language-specific functional mappings of vowel sounds. J Acoust Soc Am 104:3568–3582
Ladefoged P 1975 A course on phonetics. Harcourt Brace Jovanovich, New York
Landau B, Gleitman L 1985 Language and experience—evidence from a blind child. Harvard University Press, Cambridge, MA

Lenneberg EH 1967 Biological foundations of language. Wiley, New York
Levelt WJM 1989 Speaking: from intention to articulation. MIT Press, Cambridge, MA
Liberman M, Prince A 1977 On stress and linguistic rhythm. Linguistic Inquiry 8:249–336
Manrique AMB, Signorini A 1983 Segmental durations and rhythm in Spanish. J Phonetics 11:117–128
Mazuka R 1996 Can a grammatical parameter be set before the first word? Prosodic contributions to early setting of a grammatical parameter. In JL Morgan, K Demuth (Eds.) Signal to syntax: Bootstrapping from speech to grammar in early acquisition. Lawrence Erlbaum, Mahwah, NJ, p 313–330
Mehler J 1981 The role of syllables in speech processing: Infant and adult data. Philos Trans R Soc 295:333–352
Mehler J, Jusczyk P, Lambertz G, Halsted N, Bertoncini J, Amiel-Tison C 1988 A precursor of language acquisition in young infants. Cognition 29:143–178
Mehler J, Dupoux E 1994 What infants know. Basil Blackwell, Cambridge
Meisel JM 1992 The acquisition of verb placement. Functional categories and V2 phenomena in language acquisition. Kluwer Academic Press, Dordrecht
Miller GA 1951 Language and communication. McGraw-Hill Book Company Inc, New York
Molfese D, Molfese V 1979 Hemisphere and stimulus differences as reflected in the cortical responses of newborn infants to speech stimuli. Develop Psychol 15:505–551
Moon C, Cooper RP, Fifer WP 1993 Two-day-olds prefer their native language. Infant Behav Dev 16:495–500
Morgan JL, Meier RP, Newport EL 1987 Structural packaging in the input to language learning: contributions of prosodic and morphological marking of phrases to the acquisition of language. Cognit Psychol 19:498–550
Morgan JL, Demuth K 1996 Signal to syntax: bootstrapping from speech to grammar in early acquisition. Lawrence Erlbaum, Mahwah NJ
Morgan JL, Saffran JR 1995 Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. Child Dev 66:911–936
Nazzi T, Bertoncini J, Mehler J 1998 Language discrimination by newborns: toward an understanding of the role of rhythm. J Exp Psychol Hum Percept Perform 24:756–766
Nespor M 1990 On the rhythm parameter in phonology. In: IM Roca (Ed.) Logical issues in language acquisition. Foris, Dordrecht, p 157–175
Nespor M, Vogel I 1986 Prosodic phonology. Foris, Dordrecht
Nespor M, Guasti MT, Christophe A 1996 Selecting word order: the rhythmic activation principle. In: U Kleinhenz (Ed.), Interfaces in phonology. Akademie Verlag, Berlin, p 1–26
Nespor M, Mehler J, Peña M 2003 On the different role of vowels and consonants in language processing and language acquisition. Lingue e Linguaggio 221–247
Newport EL, Aslin RN 2004 Learning at a distance I. Statistical learning of non-adjacent dependencies. Cognit Psychol 48:127–162
Peña M 2002 Rôle du calcul statistique dans l'acquisition du langage (Doctoral dissertation, Ecole des Hautes Etudes de Sciences Sociales, 2002)
Peña M, Bonatti LL, Nespor M, Mehler J 2002 Signal-driven computations in speech processing. Science 298:604–607
Peña M, Maki A, Kovacic D et al 2003 Sounds and silence: an optical topography study of language recognition at birth. Proc Natl Acad Sci USA 100:11702–11705
Pike KL 1945 The intonation of American English. University of Michigan Press, Ann Arbor, MI
Pinker S 1994 The language instinct: How the mind creates language. Harper Collins, New York
Port RF, Dalby J, O'Dell M 1987 Evidence for mora timing in Japanese. J Acoust Soc Am 81:1574–1585
Premack D 1971 Language in chimpanzee? Science 172:808–822

Premack D 1986 'Gavagai!' or the future history of the animal language debate. MIT Press, Cambridge, MA
Querleu D, Renard X, Versyp F, Paris-Delrue L, Crepin G 1988 Fetal hearing. Eur J Obstet Gynecol Reprod Biol 28:191–212
Ramus F, Hauser MD, Miller C, Morris D, Mehler J 2000 Language discrimination by human newborns and by cotton-top tamarin monkeys. Science 288:349–351
Ramus F, Nespor M, Mehler J 1999 Correlates of linguistic rhythm in the speech signal. Cognition 73:265–292
Saffran JR, Newport EL, Aslin RN 1996 Word segmentation: the role of distributional cues. J Mem Lang 35:606–621
Saffran JR, Aslin RN, Newport EL 1996 Statistical learning by 8-month-old infants. Science 274:1926–1928
Segalowitz SJ, Chapman JS 1980 Cerebral asymmetry for speech in neonates: a behavioral measure. Brain Lang 9:281–88
Seidenberg MS, MacDonald MC 1999 A probabilistic constraints approach to language acquisition and processing. Cognit Sci 23:569–588
Selkirk E 1984 Phonology and syntax: the relation between sound and structure. MIT Press, Cambridge, MA
Stevens KN 1998 Acoustic phonetics. MIT Press, Cambridge, MA
Villringer A, Chance B 1997 Non-invasive optical spectroscopy and imaging of human brain function. Trends Neurosci 20:435–442
Wanner E, Gleitman LR 1982 Language acquisition: the state of the art. Cambridge University Press, Cambridge
Werker JF, Tees RC 1983 Developmental changes across childhood in the perception of nonnative speech sounds. Can J Psychol 37:278–286
Werker JF, Tees RC 1984 Phonemic and phonetic factors in adult cross-language speech perception. J Acoust Soc Am 75:1866–1878
Yang C 2004 Universal grammar, statistics or both? Trends Cog Sci 8:451–456

DISCUSSION

Logothetis: What you have been describing can be mathematically represented as a series of Markov chains. I thought the recent results from Mark Hauser with tamarin monkeys give a bit of a hint as to what might be happening. They have been testing the ability of these animals to detect certain sequences, randomising the process appropriately with Markov chains. They have shown that the tamarins have the basic machinery for detecting certain transitions. All is needed is for them to fine tune it. Would this be a mechanism that appeals to you?

Mehler: What Fitch & Hauser defined were 'A' items and 'B' items. Both As and Bs were syllables. There were eight 'A' syllables and eight 'B' syllables. Using these 16 syllables they tried to teach two types of grammar to the tamarins. One is an $(AB)^d$ kind of grammar while the other is an A^nB^n grammar. While the first generates sequences of syllables like (ABAB, ABABAB, etc.), the other generates sequences like (AABB, AAABBB, etc.). Unfortunately, one has to explain to the

animal which syllables are As and which are Bs. They signalled As to the animals using a high pitched voice and the Bs using a low pitched voice. Their results suggest that while humans extract both kinds of grammars from the few examples they are given, the tamarins only extract the simpler grammar, namely, the $(AB)^n$. Perhaps the monkeys aren't paying attention to the syllables at all but only to the high and low pitch tones. They may even be computing the transition probabilities between high-pitched and low-pitched sounds. I believe that their results can be explained in this way. Of course, some further tests could be used to explore whether both populations learned the intended grammar or something else. I will not reveal at this point how adults in our lab behaved when they are given sequences like HLHLH or HHHLL neither of which is grammatically compatible with the grammars that the authors tried to teach the tamarins and their human participants. However, our own work suggests that the convergence on a grammar required more than Fitch and Hauser think. Another question is whether humans but not tamarins learn some kinds of grammars and not others. Obviously this must be true, otherwise tamarins would by now be using fully-fledged grammatical structures. In short, Fitch and Hauser haven't demonstrated that the tamarins are learning either of these grammars and even the behaviour of humans needs to be evaluated with more telling tests.

Haggard: At the end of your talk it seems you were suggesting that the consonant plays a special role as the carrier of the linguistic unit. It is interesting, but I'm not sure I understand why the consonant does this and not the vowel. In particular, could it ever be the other way round?

Mehler: We explored the phonologies of many languages spoken throughout the world. Languages with lexical roots that are characterized by vowels were not found while languages in which the sequence of consonants defines lexical roots are numerous. Indeed, in Semitic languages a sequence like *GDL* is not a word but a root that can realize itself as: *gadol*, 'big' (masculine adjective); *gdola*, 'big' (feminine adjective); *giddel*, 'he grew' (transitive verb); *gadal*, 'he grew' (intransitive verb); *higdil*, 'he magnified' (transitive verb); *magdelet*, 'magnifier' (lens), etc. Thus, *GDL* is a root whose meaning is related to the 'enlarging/growing' semantic complex. Why, we can now ask, is it not possible to find similar roots defined in terms of the sequence of vowels?

Marina Nespor and colleagues have written a paper (Nespor et al 2003) entitled 'On the different role of vowels and consonants in speech processing and language acquisition', suggesting the following: there is not much you can do with a consonant except to pronounce it or mispronounce it, while with vowels the speaker can do a lot of things among which stressing it, changing slightly its volume or the typical first and second formants to suggest a given dialectal variation, etc. In other words, we have reasons to believe that to a large extent vowels tend to influence

grammatical properties, while consonant are mostly related to the characterization of the lexical items in a language

Rumiati: Another example for me would be the *r* for me or the *r* for the Hebrew speaker. They are simply characterizing the fact that we are speaking different languages which have different phonological features.

Mehler: Even I, who have a terrible accent in every language I speak, am happy to notice that after a while my listeners tend to understand and comprehend what I say. Thus, mispronunciations tend to be located mostly in my vowels and if they also attain the consonants it must be in a systematic fashion, like in my fricatives or nasals. Yet, vowel mispronunciations have been used since ancient times to identify foreigners: the word 'shibboleth' was used to discriminate foreigners in the Bible. But as far as I know, in that example, the variation is at the initial s/sh, not in the vowels!

Logothetis: So does the fact that the Hebrew script can be written without vowels mean that it gives you much more range of expressing the vowels?

Barash: There is an interesting point here. Hebrew is extreme in the sense that there is an Ashkenazi pronunciation which was used by the Jews in Europe and the other pronunciations which were used by Jews elsewhere. They are very different. For me it is difficult for me to understand Ashkenazi Hebrew.

Diamond: I have a question about the learning of grammatical rules and the distinction between A^nB^n and (AB^n). You said that ABAB can't be (AB^n), but does it really violate the rule?

Mehler: The actual grammar that can relate to Chomsky is not strictly speaking A^nB^n: Rather it describes nested syntactic constructions as we often use in natural languages. Consider, such constructions as

The boy that my mother's cousin met yesterday fell down the stairs

Clearly 'the boy' is linked to the phrase 'fell down the stairs' while 'the mother of my cousin' did not fall down the stairs. Rather 'mother' and 'cousin' met yesterday. We link 'the boy' to the 'fall down the stairs' and the phrase 'my mother's cousin met yesterday' is another constituent that is located between two parts, one at each of the extremes of the sentence. It is these nested constructions that are being referred to. If you leave out one of the constituent phrases the sentence as a whole becomes uninterpretable. None of these properties were really tested by Fitch & Hauser when they claim that humans learn an A^nB^n grammar.

Dehaene: Can you speculate on the relationship between perceptual development in the language domain and the development of the speech production system? Is it possible that there is, very early on, a covert, internal mapping between the perception and action systems? One bit of data is that when we do functional imaging in infants at three months of age, we can now reproducibly show that Broca's area is already activated when the infants are listening to their maternal language.

Mehler: That is a critical question. Twenty-five years ago we proposed that the syllable was incomprehensible if not the 'atom' of perception. Independently, Levelt (1989) showed that the syllable also acts as an atom of production. It would be strange if the perceptual procedures carry us to representations that are neither similar nor connected to the representations used to generate speech acts. Do we have direct evidence that this is so? No. Much more evidence is needed. It is exciting to notice that methods have become available to do these kinds of experiments with very young infants. Let me illustrate this with a very simple experiment. As soon as we know that an infant has learned a dozen or so words, is it possible to show that whenever the baby listens to one of those words areas that are also active during production become activated? And do such areas become more activated when the infant listens to nonce words, i.e. to *detty* instead of *teddy* there would be less activation than when the infant listens to *mimmo* than to *mommy* as the aforementioned hypothesis should predict.

Derdikman: Do you have a suggestion for why there is such a relationship between the phonetic structure of a language and its syntax?

Mehler: We have thought a lot about this question over the past ten years. First we noticed, as predicted by Nespor & Vogel (1986) that there must be some relation between prosodic aspects and syntax. The first evidence suggesting that the conjecture might be correct was when we discovered that very young infants can discriminate utterances drawn from two different pairs of languages but not from any two pairs of languages (Mehler et al 1988, Nazzi et al 1998, Ramus et al 1999). From these studies, Ramus et al (1999) proposed that rhythm (as measured by the quantity of vowel time in the typical utterance of the language and the variability of the intervocalic intervals) predicts the infants' behaviour. More recently yet, we showed that if one plots 20 highly varied languages in a rhythmic chart there is a dividing line that separates the Head-Complement languages (as most Romance languages) from Complement-Head languages (as Japanese and Basque) (Mehler et al 2004). Research in progress suggests that there may be a much more intimate relation between the sound structure of languages and the syntax they implement.

References

Gergely G, Bekkering H, Kiraly I 2002 Rational imitation in preverbal infants. Nature 415:755

Levelt WJM 1989 Speaking: from intention to articulation. MIT Press, Cambridge, MA

Mehler J, Jusczyk P, Lambertz G, Halsted N, Bertoncini J, Amiel-Tison C 1988 A precursor of language acquisition in young infants. Cognition 29:143–178

Mehler J, Gallés NS, Nespor M 2004 Biological foundations of language: language acquisition, cues for parameter setting and the bilingual infant. In: Gazzaniga M (Ed) The new cognitive neurosciences III. MIT Press, Cambridge, MA

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

Nespor M, Vogel I 1986 Prosodic Phonology. Foris, Dordrecht

Ramus F, Nespor M, Mehler J 1999 Correlates of the linguistic rhythm in the speech signal. Cognition 73:265–292

Final general discussion

Diamond: I notice that several speakers referred to some of the pioneers of psychology, neuroscience and experimental psychology in their papers, in referring to the inseparability between percepts and actions. People such as Sherrington and Adrian have made observations of this sort. I noticed the reactions of the audience: people seemed to agree with this idea of non-separability. Yet when Nikos Logothetis asked us to try to begin to define the decision-making network, I was surprised that most people agreed that we should drop the sensory part of it—at least what some people referred to as a purely sensory part—and then consider what is left in the network to be the decision-making part of it. This contradicts the agreement with the initial proposal of inseparability between percepts and actions. In the end it may be a useless exercise to try to define the transition between sensations and actions. Nevertheless, I want to reopen that question with a thought experiment. Suppose that a stimulus *a* produces a percept also called *a*, and we ask people to give a reaction a' when they experience this. Stimulus *b* produces a reaction b'. Suppose that we can change the subject's reaction through an external device. Does coming to the opposite action affect their judgement of the stimulus that occurred before? I wouldn't be too surprised if how you react to a stimulus affects your interpretation even though the stimulus has occurred before. For example, if we see a face talking, the visual input is so salient that we are convinced that the voice comes from the face. If the voice comes from a different source we continue to attribute the voice to the speaking face, and so we reinterpret the time of sound arrival to our ears according to a decision that we have made. Decisions thus affect percepts. Should we exclude sensations from the decision-making process, or is there a seamless transition?

Rizzolatti: I think it is important to keep sensation and perception separated. Think, for example, of the McGurk effect. Individuals are presented with two syllables ('ba', 'da', 'ga') simultaneously, one in the auditory and in the other in the visual modality. When the syllable presented in one modality does not match the one presented in the other modality, the individual may perceive a syllable different from both those presented. There is no reason to doubt that both visual and auditory stimuli are correctly analysed (that is the sensation is correct), yet the percept is different. When I say that perception and action results from a common substrate, I am not talking about what happens in the retina or in the cortical representation of the whiskers. In the syllable case, what is perceived depends on the language motor areas.

Sensation is a distinct process from action, although it may be influenced by it. In contrast, perception and action share the same neural substrate.

Haggard: The executive areas of the brain structure determine the incoming afferent sensation. Even primary cortex can be preset by executive areas and multimodal areas to process stimuli in a particular way. This seems not that different in principal from the active touch idea that you control your own sensory input by movement. Except that in the case of executive control of unimodal areas you aren't using your body to control your input, you are doing it entirely internally in your brain. If you wanted to be radical, you could say that in both of these cases the cognitive brain is setting up the afferent transmission to acquire good, better, optimal information. We were talking about where decisions are made. David Sparks very nicely said that it must be made before the relevant neurons in the superior colliculus fire. I think it must be made *after* active touch: if I am carrying out active touch, or my frontal lobes are preparing my somatosensory cortex for some input, then by definition I haven't yet decided what the stimulus is, and I am still trying to improve the sensory information I have about the stimulus. That's what this descending signal means.

Scott: There is an illusion generated in a rotating room in Jim Lackners' lab. First, you stand at the side of the room and get used to the velocity of the rotating room. When you make your first movement directly in front of you your arm gets 'knocked' to the side and you feel this imaginary force on your arm. Within a few movements you move straight and no longer feel any force. If the room rotation is stopped you get 'knocked' in the opposite direction. Within a few movements it is gone. This percept is completely generated from actions and what you are expecting from the sensory periphery.

Logothetis: Your perception under these conditions is also affected. I have been in that room, and the angles do not appear to be 90 degrees any more. It is not just the motion that is changed.

Scott: The sensation of the apparent force on your limb is changing. You adjust in just a few movements, and this is only your arm movements that have created that.

Treves: I was just thinking of a class of experiments done by Edmund Rolls in which he used gustatory stimuli and fed subjects to satiety. In primary cortex there is selectivity that is not affected by satiety, but in secondary cortex there is satiety. There is a gradient along the sensory cortex of how much something that is not in the stimulus can affect things.

Diamond: We would expect those gradients to be different for different systems, animals and paradigms. It would be interesting to explore them for each perceptual experience.

Sparks: When I began my career we knew a lot more about sensory neurophysiology than we did about motor neurophysiology, certainly at a cellular level. We

could do the sensory neurophysiology on anaesthetized animals, but it is hard to record single-cell activity related to movements in a paralysed animal. The motor physiologists therefore lagged behind the sensory physiologists in terms of cellular understanding. Because of this the sensory physiology has dominated the way we study sensory systems. The point I want to make is that in sensory systems, perception and cognition are not the only endpoint of sensory processing. Neither sensation nor cognition has any adaptive value in the absence of action. The brain has evolved to translate these sensory signals into motor commands. The motor system imposes constraints on the types of sensory processing that must occur. The normal way we do sensory neurophysiology ignores all of that. We should look at the types of signal transfer mechanisms that are required to interface with the format of the motor command, and new areas of sensory neurophysiology will open up. In terms of motor physiology, I'll stick to eye movement. It is well known that the execution of an eye movement is influenced by cognitive factors. There are some things that aren't typically measured that might be more sensitive than just measuring probability or latency of movement. These are the speed and duration of the movement. If you are doing neurophysiological recordings and you have neural activity that is cognitively mediated that you think may be influencing the execution of the movement, there is an optimal time to measure it. The thing to do is remember that the saccadic system is a gated system. It is only when the omnipause neurons (OPNs) are turned off that the commands to produce a saccade can occur. The activity that is going to influence the execution phase of the movement is the activity that is present at the time the movement is executed.

Also, the saccadic system has properties that can be used to assess the presence and magnitude of cognitive influences. One is that the superior colliculus has a map and its retinal and auditory inputs can activate different parts of the map simultaneously. If this occurs, the system does a vector average. It is possible to demonstrate the presence of a cognitive input using this feature of the motor circuitry. Gold and colleagues have done experiments in which a region of the brain that produces a saccadic eye movement was stimulated and looked at the development of cognitive influences by studying the trajectory of the movement. As the signal increases the stimulation-evoked movement will deviate from the control trajectory to an intermediate trajectory. This will build up in time. This is a sensitive way to assay the presence and magnitude of these cognitive variables. If you present a noise burst and look at an acoustically induced saccade, often they have a curved trajectory. Van Opstal and colleagues suggested that this was because the azimuth and elevation cues are quite different (Frens & Van Opstal 1995). It is the time and intensity of interaural differences that code information about the azimuth, but it is the spectral cues dependent on high frequency input that give elevation cues. They speculated that there is a delayed vertical component because processing the spectral cues takes longer. When they manipulated the frequency of the noise burst they

could vary the amplitude of the vertical component. These mapping properties can be used as sensitive measures of the presence and amplitude of cognitive influences.

Gold: The idea was using the effect of this vector average as an assay. The microstimulation produces an eye movement of a known vector. If it evokes an intermediate trajectory this can be used as an assay of the other activity, which is what we think of as this transformed sensory variable into motor coordinates.

Schall: There are two points I would like to make. First, it seems that this word 'decision' is being used frequently, perhaps carelessly, and out of context. Sometimes monkeys and people are faced with alternatives and choose between them for the purposes of achieving a goal. The word 'choice' can explain this kind of behaviour. I believe the word 'decision' should be reserved for those cases when there is real deliberation and the consequences are higher and more ambiguous. This is what competent humans do and are held responsible for. I think it is fair for us to ask whether macaque monkeys in physiology laboratories are ever deliberate? Even when the random dots produce 2% motion strength, are the monkeys deliberating? Perhaps not. We are certainly studying processes related to choice behaviour, but we need to be careful before we say that this is how decisions are made. My second point is that the title of this meeting is 'Percept, decision and action'. The claim is that there is nothing in the middle. There is sensation, the brain sorts it out, and then there's the mapping to the action. The complexity of our behaviour comes from our ability to map arbitrary responses onto given stimuli, but it is this mapping where all the action is taking place. Looking for a discrete decision stage distinct from the sensory representation and the motor preparation may be a fool's errand.

Derdikman: Related to your last comment, I believe that we make the mistake of assigning a decision process where it is not appropriate because we are so familiar with our own language. We have the term 'decision'. Every time we make a decision we can also be thinking of ourselves knowing that we are making a decision. We are very reflective about the things we do. However, monkeys are much less reflective. It could be that we are actually trying to impose the term 'decision' that is so familiar to us on the other species, where there is perhaps no such thing as decision making in the sense we use it as human beings.

Albright: Is accumulating information different from deliberating? We know that if it takes more time we are accumulating information to make the choice. Is that qualitatively different from what you are calling 'decision'?

Schall: I will claim that it is. It is possible to choose in the sense of acting in the context of alternatives, even when they are vague, in a more automatic sense, for example ordering a meal at a favourite restaurant. But deliberating about complex decisions, like ordering a meal at an unusual restaurant, cannot be done while you are doing something else. Deliberation entails other cognitive processes such as working memory that we know requires dedicated resources.

Albright: You do accumulate information, though. You read through the things that are on the menu; you build up something to base your choice on.

Schall: That is a natural way to think of it, but we are not guaranteed that this is the mechanism that holds for decisions such as those made by political leaders contemplating war, for example.

Barash: My bias is to think that there are intermediate states between visual and motor responses in eye movement. With regard to decisions, there are choices that monkeys make that are more automatic and less automatic.

Hasson: Humans, more than any other primate, are first and foremost social creatures. Therefore our decision to perform a certain act should be appropriate to each given social context. Perhaps the mirror system is a prime example of a system that was designed for shaping our social behaviours. This system is designed to adjust our behaviour by learning from what other people are doing. Moreover, as Rizzolatti showed in his talk, this system is highly sensitive to contextual cues. So one can conceive of the mirror system as a decision making device that intends to directly link our perception to our actions in this world.

Harris: Perhaps a way to think about a distinction between choice and decision is that the choice is focused on the stimuli in the external world, whereas a decision involves reflecting on your own action and the consequences of it.

Wolpert: We're just going to make this into a discussion of consciousness, and get stuck there. It seems like we are almost making decisions into a conscious internal discussion, and we'll then be stuck with all the same problems.

Schall: But that is the decision-making people care most about. It may be that the mechanisms of the brain are the same when we decide who to marry as when we select a soft-drink from a vending machine, but we cannot assume this.

Wolpert: If we are looking for the neural correlate of decision making, we know that one consequence of a decision is a motor act, so it is very hard to dissociate a motor act from neural correlates of decision. How are you going to do this neurophysiologically?

Schall: I decided to come to this meeting months ago, but I didn't come until two days ago. Clearly, we can choose in advance. Monkeys can too, although they probably cannot plan in advance in a manner as complicated as we.

Wolpert: Squirrels hide nuts: is this planning in advance? But you don't believe that they have thought about this.

Ditterich: Doesn't it happen relatively often that we are facing a difficult problem when we say that we will sleep on it? Then we wake up and we have made up our minds. What has happened? Was it some kind of automatic process or were you deliberating what you should do in that situation?

Schall: We say we make or take decisions, but we don't really. If it is really complicated, we say we can't make up our mind. We don't have access to how we make our decisions. They happen to us.

Diamond: That's true for every brain process, isn't it?

Wolpert: Ben Libet says we just get informed of our decisions after the motor system has made them.

Haggard: These were interesting studies in the 1980s which other people including myself have followed up (Libet et al 1983, Haggard & Eimer 1999). These studies are concerned with concepts of voluntary action independent of a stimulus. If you ask people when they first experienced the intention to make an action, on average they experience intentions a couple of hundred milliseconds before action. But of course the brain has begun to prepare the frontal motor potentials around a second before. There is a long period where your brain knows you are going to move when you don't. Philosophers love this. We need to distinguish carefully, though, between the concept of internally generated actions, which are only very remotely connected to a specific stimulus, and the sorts of situations which are more in focus at this meeting, where there is a set of stimuli and a set of responses, and perhaps a rather open relationship between them. I'm inclined to agree with Jeff Schall: where we are thinking about a mapping and can see a clear feed-forward link between stimulus and response, then decision and deliberation are perhaps not the way to think about it. We want to think more about perceptual categorization and feature extraction. At what point do we move from just mapping into real decisions? The words that seem to me to be relevant are things like induction, hypothesis making, somehow going beyond the information that is immediately present in the stimulus. How do we do this? The work of Jerry Fodor is probably relevant here. He has done some important philosophy of cognitive science. He envisaged the input and output sections of the human mind as being modular, feedforward. Then there is a non-modular central soup in the middle. He claimed that these work en bloc. They formed a very general representation of a whole series of beliefs all of which will influence the gap between the output of the sensory modules and the start of the motor modules. His view was that this Quinean inter-related property of our central representations makes them intractable to science. I am confident now that this is wrong. For example, neuroscientific experiments on the brain basis of context effects show us how these intermediate stages operate.

Rizzolatti: I think there are a lot of data from neuroscience, from the classical work of Mouncastle on the parietal lobe (e.g. Mountcastle et al 1975) to the discovery of mirror neurons—that proves that Fodor is wrong: the mysterious something between sensory and motor doesn't exist. I think we have to make realistic, neurophysiological hypotheses on higher-order cognitive processes, not to think of them as something not amenable to the scientific enquiry.

Haggard: Fodor gave up at the central point and said that we can't be scientists here. But I think we can and we should, even in the frontal cortex.

Treves: I like your characterization of complex decisions as deliberations. I would suggest a kind of Alan Turing view: maybe we should not criminalize monkeys for making simple decisions, but in human decisions, the things we are *not* so interested in are precisely those which we *can* describe. The Turing idea would be that you can provide a mathematical description of a phenomenon, like the beautiful description from Daniel Wolpert's talk. This is not what we would call a deliberation. A deliberation is something that we have difficulty describing mathematically. These are the challenges that we should address: to develop descriptions of mathematically intractable *deliberation-making*.

Diamond: Almost every sensory cortical area projects through layer 5 into motor centres. Sensory cortical areas can have a direct influence on complicated decisions.

Krubitzer: They also have very strong projections to the thalamus, which we haven't discussed. The thalamus is a very quick and powerful way of modulating incoming sensory stimuli from S1 through different levels of the cortex. The thalamus has massive input from the cortex as well as from sensory receptors. Psychophysical experiments show that detection levels are modified rapidly by what occurred prior to that. We could simply be modifying the ratio of sensory inputs coming in through the thalamus. We talk about decisions as unitary phenomena when in reality they might not be.

Derdikman: Two comments. First, I have one for Jeff Schall. Think of a huge crowd in the arena at ancient Rome: who makes the decision about whether to kill the fallen gladiator or not? It seems improbable to assume that there was a single person who was making the decision about the fate of the gladiator. Second, for Mathew Diamond, an experiment comes to mind. It is an old experiment by Held and Hein. Two kittens are sitting in two baskets. The first can walk, while the second is moved by the first kitten. Their perception of the world is totally different, although both of them have had exactly the same sensory experience.

Dehaene: As humans, we have a fairly clear, introspection of when we engage in conscious decision or deliberation, and when we do not. To track decision making in the human brain, one suggestion would be to capitalize on this distinction which is available to humans. It seems to me that at this meeting, we would have benefited from a closer examination of the neuropsychological literature, which is very clear in some respects. Consider for instance the 'alien hand' syndrome: some patients declare 'my hand is moving, but I am not in command of that action'. There are also other paradigms that allow examination of this distinction in normal subjects. I am reminded of a simple experiment by Marcus Raichle which examined the neural bases of automatization of behaviour. If you are asked to generate a verb in response to a noun, the first time you do this you have to go through a process of deliberation. You are searching for the appropriate verb for that noun. If you do that 10 times with the same list of nouns, however, then you automatize

the process of associating verbs to nouns. It is possible to image the brain activation contrasting these two states. Many areas have the same level of activity, but the parieto-fronto-cingulate network changes drastically with activity reducing when the process is automatized. It seems to me that the bulk of evidence points to a crucial role of long-distance parieto-fronto-cingulate networks in conscious decision making, as stressed by my colleagues and I in the 'global neuronal workspace' model (Dehaene & Naccache 2001, Dehaene & Changeux 2004).

References

Dehaene S, Changeux JP 2004 Neural mechanisms for access to consciousness. In M. Gazzaniga (Ed.) The cognitive neurosciences, 3rd edition. Norton, New York, Vol 82, p 1145–1157

Dehaene S, Naccache L 2001 Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. Cognition 79:1–37

Frens MA, Van Opstal AJ 1995 A quantitative study of auditory-evoked saccadic eye movements in two dimensions. Exp Brain Res 107:102–117

Haggard P, Eimer M 1999 On the relation between brain potentials and the awareness of voluntary movements. Exp Brain Res 126:128–133

Libet B, Gleason CA, Wright EW, Pearl DK 1983 Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. Brain 106:623–642

Mountcastle VB, Lynch JC, Georgopoulos A, Sakata H, Acuna C 1975 Posterior parietal association cortex of the monkey: command functions for operations within extrapersonal space. J Neurophysiol 38:871–908

Sigman M, Dehaene S 2005 Parsing a cognitive task: a characterization of the mind's bottleneck. PLoS: Biology 3:e37