LANGUAGE AND COGNITIVE PROCESSES, 1997, *12* (5/6), 585–612

Reflections on Phonological Bootstrapping: Its Role for Lexical and Syntactic Acquisition

Anne Christophe

LSCP, EHESS-CNRS, Paris, France and MRC-CDU, London, UK

Teresa Guasti

DIPSCO, Fondazione Scientifica San Raffaele, Milan, Italy

Marina Nespor

HIL and University of Amsterdam, Amsterdam, The Netherlands

Emmanuel Dupoux and Brit Van Ooyen

LSCP, EHESS-CNRS, Paris, France

"Phonological bootstrapping" is the hypothesis that a purely phonological analysis of the speech signal may allow infants to start acquiring the lexicon and syntax of their native language (Morgan & Demuth, 1996a). To assess this hypothesis, a first step is to estimate how much information is provided by a phonological analysis of the speech input conducted in the absence of any prior (language-specific) knowledge in other domains such as syntax or semantics.

We first review existing work on how babies may start acquiring a lexicon by relying on distributional regularities, phonotactics, typical word shape and prosodic boundary cues. Taken together, these sources of information may enable babies to learn the sound pattern of a reasonable number of the words in their native language. We then focus on syntax acquisition and discuss how babies may set one of the major structural syntactic parameters, the head direction parameter, by listening to prominence within phonological phrases and before they possess any words. Next, we discuss how babies may hope to

Requests for reprints should be addressed to Anne Christophe, LSCP, MSH, 54 Boulevard Raspail, 75270 Paris Cedex 06, France. E-mail: anne©lscp.ehess.fr

The work presented here was supported by a grant from the Direction des Recherches et de la Technologie to the first author (No. 8780844). The Human Frontiers Scientific Program and the Human Capital and Mobility Project also contributed to this research. We would like to thank Jacques Mehler and John Morton for their help in shaping this research. Many parts of this paper derive from the discussions held in Sperlonga at the "Cognitive Models of Speech Processing" conference.

acquire function words early, and how this knowledge would help lexical segmentation and acquisition, as well as syntactic analysis and acquisition.

We then present a model of phonological bootstrapping of the lexicon and syntax that helps us to illustrate the congruence between problems. Some sources of information appear to be useful for more than one purpose; for example, phonological phrases and function words may help lexical segmentation as well as segmentation into syntactic phrases and labelling (NP, VP, etc.). Although our model derives directly from our reflection on acquisition, we argue that it may also be adequate as a model of adult speech processing. Since adults allow a greater variety of experimental paradigms, an advantage of our approach is that specific hypotheses can be tested on both populations. We illustrate this aspect in the final section of the paper, where we present the results of an adult experiment which indicates that prosodic boundaries and function words play an important role in continuous speech processing.

INTRODUCTION

Considering the difficulties psycholinguists are having in trying to model speech comprehension in adults, who have full mastery of their mother tongue, it seems an almost hopeless quest to worry about language acquisition. Even endowing children with a rich innate knowledge, namely Universal Grammar, which includes the principles that are shared by all the languages of the world and the parameters that fix their variations (Chomsky, 1981), does not make the task of language acquisition a trivial one. Children still have to learn all the language-specific aspects of their mother tongue, including its words, its sound patterns and its grammatical parameters. Some researchers believe that there are major bootstrapping problems for acquiring these language-specific aspects (mainly for the lexicon and syntax; see, e.g. Gleitman & Wanner, 1982; Morgan & Demuth, 1996b).

To give an example of this bootstrapping problem in the acquisition of language, researchers working on the acquisition of syntax have generally assumed that the input to the syntactic analyser is a string of words. But getting at the string of words is not so easily done. Most models of lexical access in adults (with full knowledge of the lexicon) use syntactic and semantic information to solve the many segmentation ambiguities that arise when one considers speech as an uninterrupted string of phonemes. Babies may solve circularities such as this one by relying on an independent source of information such as prosody—that is, the intonation of speech (see, e.g. Gleitman, 1990; Gleitman & Wanner, 1982; Morgan, 1986). In the introduction to their recent book on language acquisition, Morgan and Demuth (1996a) introduce the term "phonological bootstrapping" to express the more general idea that a purely phonological analysis of speech may give babies some information about the structure of their language.

In this paper, we focus on the very beginning of language acquisition, and consider processes that may occur during the first year of life. More specifically, we discuss how babies may start building a lexicon, and how they may start acquiring rudiments of syntax (i.e. building the skeleton of a syntactic tree). For each of these problems, we assess the phonological bootstrapping hypothesis; that is, we examine how much a baby can hope to learn by relying solely on a phonological analysis of the speech signal, as well as the available experimental evidence that babies may use the proposed sources of information. We end this review by offering a model of the first stages of speech processing that integrates all the phonological information and processes that could be used by babies when acquiring their lexicon and syntax.

PHONOLOGICAL BOOTSTRAPPING OF LEXICAL ACQUISITION

At some point during the process of language acquisition, babies have to learn the words of their language. It is well known that the mapping between sound and meaning is by and large arbitrary: the same concept is encoded through very different sounds in different languages (e.g. "dog", "chien", "hund"); two similar concepts do not usually sound alike (e.g. "wristwatch", "clock"); and two similar sounding words can refer to widely different meanings (e.g. "sat" and "fat"). This very arbitrariness makes language special, in that the mapping from sound to meaning can be performed only at the level of words (or morphemes). In what follows, we refer to the entities in the mental lexicon as words, although they may be smaller than written words in some cases (i.e. morphemes) and bigger in others (i.e. phrasal expressions). A mental dictionary or lexicon embodies these arbitrary relationships between sound and meaning. It comprises semantic representations for words, a way to access these representations when listening to speech, and a way to pronounce the words when speaking (for literate people, additional mappings to written words are necessary).

The mapping of the acoustic input onto words appears to be particularly tricky for two reasons. First, speech is continuous—there are no clear markers to separate one word from the next. Secondly, the acoustic waveforms for the same word pronounced by two different speakers, or even in two different ways (e.g. whispered or shouted), are very different. In other words, the mapping between sound and meaning is a many-to-one mapping: many different sound shapes correspond to the same word. These two problems—of segmentation and of categorisation—are tricky enough when one tries to develop a model of how sound-to-meaning mapping functions in the adult speaker of a language. But they become even more so when one thinks about how a baby could acquire this mapping. For instance, in many contemporary models of lexical access, rather than first segmenting the speech stream into bits and then identifying the bits as words, adult speakers of a language are supposed to identify known words directly from the speech stream. Such a strategy is not available to babies, whose task is precisely to learn about the words of their language.

We believe that learning the sound-to-meaning mapping becomes much easier if one postulates an intermediate representation that contains abstract word forms. The mapping would then proceed in two steps: first, identify the word forms in the speech stream, then map these word forms onto their meanings. This distinction does not make much of a difference for a model of adult speech processing. Most models of lexical access in adults try to model the identification of word forms from the acoustic signal without direct reference to semantic representation, except as a means of distinguishing between alternative segmentations (either in two steps or by implementing feedback from semantic representations to word units; see, however, the exception of Gaskell and Marslen-Wilson's model in this volume). However, adding an intermediate representation between the acoustic signal and semantic representations makes a big difference when one thinks about the acquisition of a lexicon: it means that babies could learn about the two mappings independently.

The hypothesis is, therefore, that babies may bootstrap lexical acquisition through a purely phonological analysis of the speech input, that would allow them to discover the word forms of their language (or at least a significant number of them). They would learn about the mapping between word forms and their meanings in a second step. What may this purely phonological analysis consist of? Four sources of information have been identified that could be used by infants when trying to find word forms in the continuous speech stream: distributional regularity, phonotactics, typical word shape and prosodic boundary cues.

As formulated by Brent and Cartwright (1996), "*distributional regularity* refers to the intuition that sound sequences that occur frequently and in a variety of contexts are better candidates for the lexicon than those that occur rarely and in few contexts". Brent and Cartwright (1996) formalised this intuition and offered algorithms to test the feasibility of such a proposal. They showed, using a transcript of speech directed to children, that an algorithm based on distributional regularity can discover about 40% of the words in their right place with an accuracy of 47%, a performance significantly better than that of an algorithm which places word boundaries at random (but knows how many boundaries should be placed). Is there any evidence that infants pay attention to distributional regularity? Goodsitt, Morgan and Kuhl (1993) devised an experiment to test infants' sensitivity to co-occurrences between syllables. They showed that 8-month-old infants were better at processing trisyllabic strings where a pair of syllables consistently co-occurred (such as "gakoti" and "tigako", which may be represented with only two units, "gako" and "ti") than trisyllabic strings where all syllables behaved independently (such as "gakoti" and "tikoga", which have to be represented with three independent units, "ga", "ko" and "ti").

"Phonotactics" refers to the constraints on the occurrence of sounds within words and sentences. Certain sounds or combinations of sounds occur only at word edges, others only occur word-internally, or not at all. English phonotactic constraints are mostly expressed in terms of the consonant clusters that may appear at the beginning and end of syllables; thus, a string of phonemes such as [tʃstr] necessarily contains a word boundary between the [tʃ] and the [s] (as in "much stress" for instance), because it is not a possible word-internal cluster, and all other segmentations give rise to illegal word beginnings or endings. Other examples of phonotactic constraints include the vowel harmony phenomena that operate in languages such as Turkish (there is a high probability of encountering a word boundary between two vowels that do not share a given feature). It can be shown that by using constraints on the consonant clusters of English, one can improve the performance of algorithms which try to locate words without a lexicon (Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997). With respect to babies, Friederici and Wessels (1993) showed that 9-month-old Dutch infants listened longer to syllables that respected the phonotactic constraints of their language (namely, syllables that had possible onset and coda consonant clusters, such as "BRef" and "muRT", versus syllables that had impossible clusters, such as "feBR" and "RTum"). Similarly, Jusczyk et al. (1993b) used lists of low-frequency English and Dutch words that differed only in their phonotactic patterns (e.g. Dutch but not English allows the [vl] word-initial clusters such as in "vlammend"; English but not Dutch allows a word-final voiced consonant such as in "hubbub"). American babies of 9 months of age preferred to listen to the English than to the Dutch lists, while Dutch infants showed the reverse preference. These results show that infants know about the phonotactic constraints of their native language from the age of 9 months, and it is therefore conceivable that infants use this knowledge to help word segmentation.

A well-studied example of how *typical word shape* may help lexical acquisition rests on the fact that, in English, content words predominantly start with syllables containing a full vowel (Cutler & Carter, 1987). Hence, it is useful in English to posit or favour a word boundary before a strong syllable and it has been shown that this is precisely what English adults do (see, e.g. Cutler & Norris, 1988; Norris, McQueen, & Cutler, 1995). Babies could also exploit this regularity of English, provided they have a way to learn about it (Cutler, 1996). Jusczyk, Cutler and Redanz (1993a) showed that 9-month-old American infants chose to listen longer to lists of bisyllabic words that exemplify the most common pattern in English, namely strong–weak words rather than weak–strong words. Moreover, recent experimental work with American infants aged 8–12 months suggests that they actually use this regularity of English when processing sentences. Newsome and Jusczyk (1995) exposed 7½-month-old infants to speech passages in which a particular word (e.g. "kingdom") occurred in each sentence. After this familiarisation phase, the infants were presented with the same word repeated many times, or with a different word (e.g. "candle"). In the preferential looking procedure, infants chose to listen longer to the word they had been exposed to during the familiarisation phase. However, this worked with strong–weak words (such as "kingdom" and "candle"), but not with weak–strong words (such as "guitar"). This is an indication that young English-speaking babies, just like English-speaking adults, find it easier to extract strong–weak than weak–strong words from the continuous speech stream (see also Childers & Echols, 1996; Morgan, 1996).

In addition to these processes that rely mostly on a segmental representation of speech, we have proposed that infants could draw on a fourth source of information, namely the prosody of speech, its melody and rhythm (Christophe, 1993; Christophe & Dupoux, 1996; Christophe, Dupoux, & Mehler, 1993). Syllables and words in a sentence do not occur with a metronomical regularity, nor are they pronounced in a monotone, robot-like voice: they are grouped into prosodic units. Listeners perceive these groupings, and it has been shown that they do so on the basis of an analysis of the incoming acoustic signal, rather than through a reconstruction process that involves lexical and syntactic processes. Thus, de Pijper and Sanderman (1994) showed that listeners made similar judgements about the depth of boundaries between pairs of words when they listened to natural sentences, and to the same sentences rendered unintelligible while preserving their prosody (through a process involving resynthesis with fixed formant values). The prosodic units perceived by listeners corresponded roughly to phonological phrases as defined by theories of prosodic phonology (modulo some performance constraints; see, e.g. Bailly, 1989; Delais-Roussarie, 1995; Dirksen, 1992; Pasdeloup, 1990). They typically contain one or two content words together with their clitics and, at least in some languages, they tend to contain less than about seven syllables (possibly depending on speech rate; e.g. Bailly, 1989; Dirksen, 1992). Of the prosodic phenomena that have been suggested to signal phonological phrases in at least some languages, one can include pre-boundary lengthening (e.g. Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992) and the fact that each phonological phrase may contain one or more pitch accents and may end in a boundary tone (Hayes & Lahiri, 1991).

Since prosodic phrase boundaries are marked, those word boundaries that coincide with prosodic phrase boundaries can be perceived by listeners without reference to the lexicon. Could babies perform such a prosodic analysis? Christophe, Dupoux, Bertoncini and Mehler (1994) showed that French newborns are able to discriminate bisyllabic strings that differ only in the presence or absence of a phonological phrase boundary. This suggests that phonological phrase boundary information is perceptible by very young infants and may therefore be used for segmentation purposes. More directly, Hirsh-Pasek et al. (1987) have observed sensitivity to intonational phrase boundaries (roughly corresponding to clause boundaries) in babies listening to continuous speech: In the preferential listening procedure, 6- and 9-month-old babies prefer to listen to speech interrupted by 1 sec pauses at intonational phrase boundaries rather than within an intonational phrase. Using the same procedure, Gerken, Jusczyk and Mandel (1994) noted that 9-month-olds also react to phonological phrase boundaries (specifically a boundary between a full NP subject and a verb).

To summarise, the sources of information just reviewed all appear to be plausible candidates to help babies acquire a phonological lexicon. The potential usefulness of each of the proposed processes has been demonstrated, and at least preliminary experimental evidence on infants suggests that the information necessary for them to work seems to be available before the end of the first year of life. This means that these procedures may be available in time for the start of lexical acquisition (the available studies suggest that babies may start linking word forms to meanings at about the age of 1 year at the earliest; see, e.g. Oviatt, 1980; Thomas et al., 1981). In the next section, we turn to the problem of performing a rudimentary syntactic analysis of the sentences perceived by babies.

HOW TO SET THE HEAD-DIRECTION PARAMETER BEFORE THE FIRST WORD

Traditionally, the acquisition of syntax is supposed to begin when children possess a sufficient number of words, together with their meanings. Children are supposed to start analysing sentences by parsing them into words, and then figuring out the structural relationships between the words from the meaning of the words and sentences, and from distributional analyses about the co-occurrence of words (see, e.g. Radford, 1990). However, we have just seen that merely finding the words in the continuous speech stream is not so easy, and assigning meaning to word forms is not easy either (Gleitman, 1990). Some people have criticised this approach, that syntax has to be learnt by analysing strings of words, arguing that it can give rise to paradoxical situations (Mazuka, 1996). In this section, we examine the possibility that babies may exploit a link between the phonology and the abstract structural properties of a language, which involves the order of heads relative to their complements.

In some languages, such as English, Spanish, Greek and French, main clauses typically precede subordinate clauses. In others, such as Turkish, Bengali and Japanese, the reverse relative order of main and subordinate clauses is found. The first type of languages is also characterised by generally having complements following the head, where a head can be a verb, a noun, a preposition or an adjective. In the second type of languages, the complements generally precede the head. We can summarise that by stating that languages such as English expand to the right, whereas languages such as Turkish expand to the left. Examples from English and Turkish are as follows:

I want that Marina receives this message (main–subordinate)	Marina'nIn bu mesagI almasInI istiyorum Marina-gen this message-acc receive-3sg-acc want-lsg (subordinate–main) "I want that Marina receives this message"
I wrote the book (verb–complement)	KitabI yazdIm the book-acc. write-lsg-past (complement–verb) "I wrote the book"
The writing of the books (noun–complement)	sehrin Iklimi the city-gen. climate (complement–noun) "the climate of the city"
for you (preposition–complement)	senin için you for (complement–postposition) "for you"

Discovering how words are organised within phrases and phrases within sentences is crucial for children, because it is a prerequisite for any syntactic analysis. In the principle and parameter model (Chomsky, 1981), this structural property of languages is expressed by the head-direction parameter that bears on the X-bar schema, which constrains the way a possible phrase structure is built. Within the X-bar schema, complements are sisters of a head (X): they appear on the left in left-recursive languages and on the right in right-recursive languages. On this view, the only thing babies have to do is to decide whether, in their language, complements generally precede their heads or the reverse. A tacit assumption in the literature has been that children first segment the speech stream into words, then learn and categorise some isolated words and, finally, set the value of the word order parameter by looking at how words are combined in the input available to them. We offer here an alternative explanation (Nespor, Guasti, & Christophe, 1996), which rests on a purely phonological analysis of the speech input (see also Mazuka, 1996).

Our proposal rests on the fact that the head-direction parameter possesses a prosodic correlate: in languages whose syntactic trees are right-branching, the rightmost word of phonological phrases is the most prominent, whereas the leftmost word of phonological phrases is the most prominent in languages whose syntactic trees are left-branching (after Nespor & Vogel, 1986). As a consequence, the location of the prominence within phonological phrases may indicate to babies how a phrase structure is to be built. Of crucial importance is that this information may be available to babies at some time during the first year of life (i.e. in time for the onset of lexical and syntactic acquisition).

How plausible is this hypothesis? One of its consequences is that the word order parameter should be set correctly very early. Can we find any evidence for this? As far as language production is concerned, it is known that when children start combining words, they hardly deviate from their target grammar order (Bloom, 1970; Brown, 1973; Pinker, 1994). This implies that the word order parameter is probably set before the age of about 2½ years. Only comprehension studies show us whether it is set earlier than that. Roberts (1983) found evidence that children of about 2 years of age exploited the word order of reversible sentences so as to infer who were the agent and patient of the action to be carried out (e.g. "can [child's name] kiss [caretaker]?"). Similarly, Hirsh-Pasek and Golinkoff (1996) report evidence that children as young as 17 months may be able to use word order information to comprehend reversible sentences: they look longer at a video illustrating the correct meaning of the sentence (e.g. "Cookie Monster is tickling Big Bird"). In sum, the available experimental evidence suggests that the parameter governing word order is set to the correct value as early as one can investigate, congruent with our hypothesis.

Next, for our hypothesis to work, we must assume that babies are sensitive to both phonological phrase boundaries and the main prominence within phonological phrases. How plausible are these two claims? We have previously presented experimental evidence that suggests that infants under 1 year of age can perceive phonological phrase boundaries in continuous speech (Gerken et al., 1994). As regards prominence information, there is no direct evidence that babies perceive which word carries most stress within a phonological phrase. However, note that it would be sufficient for them to spot the most salient syllable within phonological phrases. Although this does not systematically correlate with the direction of recursivity of the language, in most cases the most salient syllable will be towards the end in right-recursive languages, and vice versa for left-recursive languages. Mismatches may happen whenever a phonological phrase contains only one word, for instance: In such cases, the most salient syllable will be the one that carries lexical stress, and its position will depend upon the characteristics of lexical stress in this language, not upon its direction of recursivity. Do we have any experimental evidence to suggest that babies can identify the most salient syllable of a string? We know, for instance, that very young infants can discriminate between bisyllabic utterances that have stress on the first versus the second syllable (Jusczyk & Thompson, 1978; Sansavini, 1995). However, this does not mean that babies can identify the most salient syllable of a string of 6 or 7. However, it makes this hypothesis more plausible.

For our hypothesis to be taken seriously, what we really need is experimental evidence that babies can perceive prominence within phonological phrases. Experimental work involving a comparison of French and Turkish, which share many prosodic properties (both have a simple syllabic structure, lexical stress is word-final, and there are no reduced vowels) but differ for the head-direction parameter (French is head-initial while Turkish is head-final), is currently being carried out.

To summarise, we have argued that babies may be able to decide about one of the main structural properties of their language, namely the relative order of heads and complements, simply by listening to prominence within phonological phrases. This information, if babies are able to exploit it, would be available within the first year of life. Therefore, at the end of the first year of life, when we might think that babies start segmenting sentences into words, they may already possess one of the bits of knowledge necessary to start building X-bar structures. We now turn to another source of information that would be particularly useful in conjunction with the knowledge about the head-direction parameter in order to perform a rudimentary syntactic analysis, namely function words.

LEARNING ABOUT FUNCTION WORDS

Function words come at the junction between lexicon and syntax (they are the grammatical words, like determiners, auxiliaries, prepositions, etc.). They are words that have to be segmented out of the continuous speech stream and learned, but they also play a very important role in syntactic structure. In fact, babies may be able to discover function words quite early in their acquisition of language. Morgan, Shi and Allopenna (1996) propose that function words share some fundamental phonological, statistical and acoustic properties across the languages of the world, which may make them rather easy to spot by babies (e.g. in English, function words typically contain a reduced vowel, they are shorter and more central, they tend to be unstressed and monosyllabic; in Mandarin Chinese, function words are typically shorter than content words, they tend to carry neutral tone, etc.). Morgan et al. present the results of an analysis of infant-directed speech in both English and Mandarin Chinese showing that when several cues are taken together, they allow function words to be distinguished from content words with reasonable accuracy.

One of the cues suggested by Morgan et al. (1996) concerns the distribution of function words in sentences. If we assume that babies have access to phonological phrase boundaries (as we have done above for other reasons), then these distributional cues become much more powerful and may also be used even before babies are able to parse sentences into words. Because phonological phrases are constructed with reference to syntactic structure, closed-class items such as determiners, pronouns and conjunctions tend to occur at their edges (beginning or end in general, depending on the language). Thus it is a good strategy to look for closed-class items at the edges of prosodic units. Babies could exploit this regularity of languages and compile a list of the syllables that occur at the beginning and end of prosodic units, storing the most frequent syllables in a separate list, and subsequently identifying these syllables as closed-class items when encountered at the borders of a prosodic unit.

If anything like this is correct, we would expect to find an early sensitivity to closed-class items. LouAnn Gerken and her colleagues (Gerken, 1994; Gerken & McIntosh, 1993) have studied 2-year-olds who do not produce function words consistently and have shown in a variety of tasks that these words are nevertheless fully processed. But a much earlier sensitivity would be necessary if function words are to help bootstrapping syntactic and lexical acquisition. In an ERP study, Shafer, Gerken, Shucard and Shucard (1992) noted that 10- to 11-month-olds react differentially to English stories with normal function words versus stories where function words are replaced with anomalous monosyllables (e.g. /gu/). Their results favour the hypothesis that babies may start compiling a list of the function words in their language before the end of the first year of life.

How would babies exploit an early knowledge of function words? First, function words may be used in conjunction with prosodic information to perform an initial segmentation and labelling of sentences into syntactic constituents. Prosody may be a cue to syntactic structure (Gerken et al., 1994; Morgan, 1986), but in itself provides no cue to the labelling of constituents (into noun phrase, verb phrase, etc.). Thus, a phonological phrase boundary always coincides with a syntactic boundary (X''). But the match between phonological phrases and syntactic constituents is not perfect. For instance, many syntactic boundaries are not prosodically marked, and the biggest prosodic boundary in a sentence does not necessarily correspond to the highest syntactic juncture (cf. Chomsky & Halle, 1968). Nonetheless, phonological phrase bracketing furnishes a good first approximation of syntactic bracketing. For instance, in [The fast car] [arrived] [before the train], all phonological phrase boundaries coincide with left X″ boundaries and prosodic bracketing isolates the main syntactic constituents of the sentence. Obviously, children may exploit their knowledge of function words and morphemes to label the constituents demarcated by the prosody (see Gerken, 1996; Gerken, Landau, & Remez, 1990; Morgan, 1986). To do this, babies would first need to identify which function words occur in noun phrases versus verb phrases, for instance. They may be helped in that by some universal properties of language; for instance, within one intonational phrase, there is at most one VP, but there may be more than one NP (Nespor et al., 1996). Function words and morphemes that can appear in more than one phonological phrase within an intonational phrase can therefore be attributed to NPs, while those that occur only once can be attributed to VPs. Subsequently, children may be able to generalise what they have learnt to label syntactic phrases and also categorise new content words as nouns or verbs.

We have focused up to now on how babies may use knowledge of function words to help early syntactic acquisition and parsing. Knowing just the form of function words, however, may also help in discovering the forms of the content words of one's language. We mentioned above that function words tend to occur at the borders of prosodic units. As soon as a list of function words is established, babies could strip off those very frequent syllables when encountered at the borders of prosodic units and identify the rest of the string as a content word. This strategy could be called "function-word-stripping".

To summarise, we have argued that children may be able to identify the function words of their language very early, through a combination of their distributional properties relative to prosodic phrase boundaries, and of their specific phonological properties. Children may also hope to bootstrap their knowledge of which function words go with the main grammatical categories (mainly nouns and verbs) through a distributional analysis of sentences. Although this is fairly speculative at this point, we wish to emphasise the many benefits that babies might draw from such knowledge: on the one hand, knowing what syllables are function words may allow them to refine their learning of content words within prosodic phrases; on the other, knowledge of both prosodic phrase boundaries and of function words may allow babies to perform a first syntactic bracketing and labelling. In addition, we claim that at the same time as this knowledge is available, babies may already know how to build the X-bar trees for their language.

Taken together, this implies that babies may be able to make a rudimentary syntactic analysis of a sentence even when they do not know any of the content words of that sentence. Possessing such knowledge early, at an age when babies still know very few content words, would be very useful in many respects. For instance, babies could use the syntactic information to constrain their acquisition of the meaning of the words—they may be able to identify which word functions as the verb, which noun is subject and which is object of that verb. Such knowledge would make it much easier to figure out the meaning of the individual words from the scene created by the meaning of the sentence (Gleitman, 1990).

A MODEL OF PHONOLOGICAL BOOTSTRAPPING, AND OF THE EARLY STAGES OF SPEECH PROCESSING IN ADULTS

We have so far reviewed potential phonological bootstrapping solutions to several acquisition problems. We now review the advantages of considering these problems together, and emphasise the convergence between the solutions proposed. Throughout our exposition, we have repeatedly come back to the same processes and representations. Thus, exploiting the distributional properties of language seems useful at many levels of structure: to discover phonotactic regularities, to find function words (in conjunction with prosodic boundary information) and to find content words. Similarly, perceiving phonological phrases in the acoustic signal would help babies discover content and function words, set the head-direction parameter (in conjunction with prominence information) and perform a rudimentary syntactic analysis of sentences (in conjunction with function word information). It seems to us that the plausibility of a given mechanism or representation becomes greater when its benefits to children are shown to be so widespread. At the very least, it motivates us more to find definite proof of such processes or representations.

In addition, we would like to view the proposed processes in light of the adult speech processing system. The rationale behind this move is that if a given source of information is reliable enough to be able to bootstrap acquisition, there is no reason why adults would not also make use of it when processing sentences. In addition to the fact that it would be inefficient to ignore useful information, the adult language-processing system has developed from the baby's processing system in interaction with the linguistic environment. This does not necessarily imply that any single process that was used during acquisition is still used by adults. It may well be that some processes become obsolete as full knowledge of the native language allows more efficient procedures to develop. Nevertheless, it seems to us a good idea to start with the hypothesis that infants and adults share the architecture of their language-processing systems, the former to acquire language and the latter to process it (Mehler, Dupoux, & Segui, 1990). In line with this, Fig. 1 represents the early stages of speech processing both in adults and in infants.

As in many models of speech processing (with the exception of Klatt, 1979; Marslen-Wilson & Warren, 1994), we postulate the existence of a pre-lexical level of representation that mediates between acoustic–phonetic processes and lexical processes. Our specific motivation for this is that all the processes identified in the top half of Fig. 1 rely on regularities in the language that cannot be discovered on a raw acoustic representation: the input to the mechanisms exploiting this type of regularity has to be in the form of categorised units such as phonemes, syllables, demi-syllables, etc., which are normalised for speaker and rate of speech.[1]

In addition, we assume that prosodic domains are represented at the same pre-lexical level of representation. Thus, prosodic boundary information is supposed to be available to all lexical processes. As a consequence, it is expected to constrain lexical access in adults: lexical access should start and stop at prosodic boundaries (as well as in other places). In the course of acquisition, prosodic boundaries will also bootstrap mechanisms based on the regularities of language, such as phonotactics or typical word shape. Such a claim is based mainly on *a priori* reasoning: prosodic information *would* be extremely useful in this place, and it is very likely to be available. However, it remains to be tested experimentally. Although it may be hard to find experimental evidence that prosodic information allows infants to discover the phonotactics of their native language, it is relatively straightforward to design experiments to test adults' use of prosodic information to constrain lexical access (see next section for specific proposals); we also have some information about babies' perception of prosodic boundaries (Gerken et al., 1994) and it is conceivable to find out more. We believe that our claim would be more plausible if we could show that adults use prosodic boundary information in the course of lexical access *and* that infants are sensitive to the same information.

A third claim made by the model is that the pre-lexical representation is to a large extent language-specific. That is, the phonemes, syllables and features linked to the syllabic frames which it contains represent the acoustic information conveyed in speech in a language-specific way: some of the irrelevant variation is no longer represented at this level. Our reason for this is that a language-specific representation would make the exploitation of language regularities much easier. Let us illustrate this with an example:

[1] The exact nature of this representation is of little importance to the present discussion; it could also consist of features linked to syllabic or skeletal positions (see Gaskell and Marslen-Wilson, this volume).

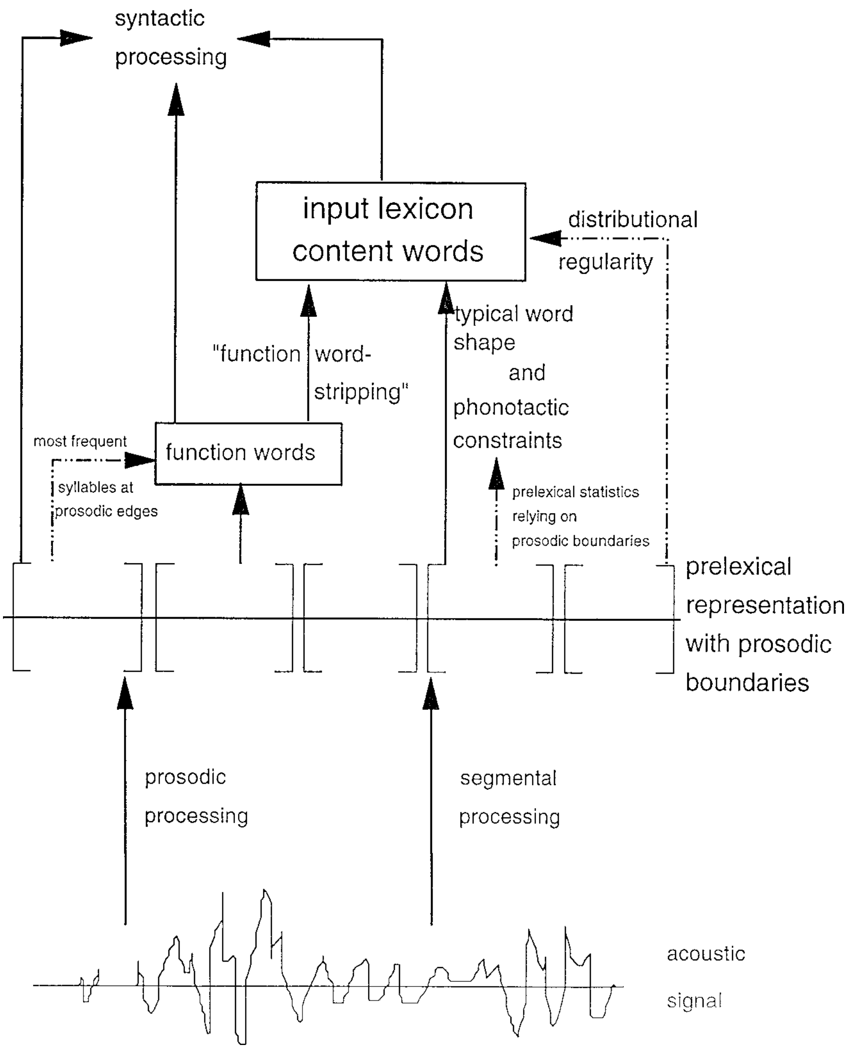

FIG. 1. A model of the early stages of speech processing: solid lines and boxes represent processes in the adult; dotted lines and boxes represent processes specifically at work during acquisition.

Spanish has a five-vowel system, so that the only front mid-vowel is /e/; variability in production is such that some of these /e/ vowels will in fact be far enough from the /e/ prototype to be perceived as, for instance, /ɛ/ by speakers of another language (e.g. Catalan). If babies relied on a universal representation of the vowel space (such as the one that Kuhl, 1992, proposed to be initially available to infants), they would represent differently two vowels that are in fact the same and this may disrupt the search for regularities. In parallel, there is evidence that 6-month-old babies already organise their vowel space, like adults do, from their linguistic environment (Kuhl et al., 1992). The claim is, therefore, that this language-specific knowledge should be present in the pre-lexical representation. To summarise, our claims about the pre-lexical representation are: (1) that it exists, (2) that it is prosodically segmented, and (3) that it is language-specific.[2]

The top half of the model in Fig. 1 features the exploitation of the regularities of the language (distributional regularities, phonotactics, frequent word patterns, special use of function words; note that the boxes and arrows are not meant to depict actual processes, but rather are an indication of what kind of information is used, and what levels it connects). The claim here is, therefore, that all these sources of information are exploited simultaneously in the process of accessing and constructing patterns in the input lexicon from the pre-lexical representation. One example of the successful integration of language-specific information into a lexical access model is the Shortlist model developed by Dennis Norris (see Norris et al., 1995; Norris, 1994). In the Shortlist model, information about the most frequent lexical pattern for English content words (initial strong syllable) is used to constrain the patterns of activation of candidate words as the input develops over time. Not only is the language-specific information useful to the system (allowing faster convergence of words conforming to the pattern—the most frequent case), but the behaviour of the model is similar to that of native English adult speakers (see, e.g. Cutler & Norris, 1988; McQueen, Norris, & Cutler, 1994). Phonotactic constraints are similar in kind and could probably be incorporated at the same level with fruitful results. As far as acquisition is concerned, language-specific knowledge of both phonotactic regularities and typical word shape could be obtained by analysing the boundaries available (those that are prosodically marked) and by computing statistics on the words already present in the lexicon. Both

[2] Both prosodic and segmental processing (in the bottom half of the model in Fig. 1) are needed to map the acoustic signal to the pre-lexical representation. We have depicted them as operating independently from one another, mainly for the sake of simplicity. However, we have no evidence in favour of this view, and it may be that it is not possible. Thus, a vowel with a greater than normal duration may be a phonemically long vowel, a stressed vowel, or the last vowel of an intonational phrase: the same characteristic, duration, codes several things. It may be that this problem can be solved by considering simultaneously all the prosodic parameters. Thus, if the long vowel is stressed, it will also be higher-pitched; if it is at the end of an intonational phrase, it will be part of a decreasing pitch contour; and in both cases, neighbouring segments will be lengthened as well, whereas only the vowel will be lengthened if it is phonemically long. Prosodic characteristics will generally be marked by a cluster of phenomena and this may allow them to be distinguished from segmental characteristics, although both are carried by the same medium.

processes probably play a role. But the fact that infants can obtain this knowledge pre-lexically implies that these processes can be used from the very beginning of lexical acquisition.

In contrast, function-word-stripping appears to be different in kind from the other processes described, in that we have to postulate a special representation for function words. Function words and morphemes have to be listed separately (or be represented by a different population of neurons depending on the model), or be tagged with a special feature, to play a special role in lexical access. This level of representation also links directly to syntactic processing. According to our model, the very first representation available to the syntactic parser is a prosodically segmented representation where function words are filled in. Content words would be filled in as they become available from the content word lexicon. It should be possible to gather experimental data to test predictions of this model. As regards acquisition *per se*, the implication of the model is that some syntactic analysis can be performed even in the absence of clear information from the content word lexicon.

THE ROLE OF FUNCTION WORDS AND PROSODIC BOUNDARIES IN ADULTS' LEXICAL ACCESS

To test the model, we decided to focus on the issue of how adults access words in sentences. One of our aims was to design an appropriate on-line method to study this question. Until now, lexical access in sentences has been mainly studied using the cross-modal semantic priming technique. With this technique, a sentence is auditorily presented and at some point in the sentence a word appears on a screen. Subjects typically have to perform a lexical decision on the visually presented word. The semantic relationship between the visual target and a word present in the sentences is manipulated to gain insights into lexical access to the auditorily presented word (Gow & Gordon, 1995; Tabossi, Burani, & Scott, 1995). We wished to devise a task that could also be performed before lexical access is completed, so as to allow us to study the pre-lexical segmentation procedures that are of particular interest to us (e.g. prosodic boundaries, phonotactics). Note that adults' use of the strong–weak bias for English has been studied extensively, but never with on-line tasks and auditorily presented sentences simultaneously. These experiments have either featured sentences with off-line tasks (e.g. Cutler & Butterfield, 1992), or on-line tasks with short strings of nonsense syllables (e.g. Cutler & Norris, 1988).

We asked subjects to detect a target phoneme within a sentence *only if it appeared at the beginning of a word*. Subjects should find this task rather easy whenever information about this particular word boundary is readily available, and should be slowed down otherwise. We positioned the target phonemes either at the beginning of a noun following a determiner, or at the beginning of an adjective following a noun. In addition, another group of subjects was asked to detect phonemes whatever their position in the sentence (whether at the beginning of words or not) to obtain some baseline information about the speed of detection of the target phonemes independently of their position within words.

Method

Materials. Forty-four experimental sentences contained a word-initial target in a noun phrase of the form "determiner–noun–adjective" (the default ordering in French). This noun phrase was situated either at the beginning, in the middle or at the end of generally rather long sentences. For half of these sentences ($n = 22$), the target occurred at the beginning of the noun, just after the determiner (e.g. /f/ "un Fou larmoyant", "a crying fool").[3] For the other half, the target occurred at the beginning of the adjective, just after the noun (e.g. /g/ "son pas Gracieux", "his graceful walk").[4] In addition, 22 filler sentences were constructed, in which the target appeared at the beginning of a syllable but in the middle of a noun or adjective (e.g. /p/ "une raPière aiguisée", "a sharp sword"). Subjects had to respond to these targets in the generalised version of the task but not in the word-initial one. Finally, another 32 sentences did not contain the target at all. The target phonemes used in the experiment were p, t, k, b, d, g, f, s, v, j, m, n and r, all of which were present in the three subsets of sentences (target at the beginning of a word, target in the middle of a word, no target at all).

All sentences were read by a native French speaker who was naive to the aims of the experiment and who was asked to read naturally but at a rather fast speech rate. Four versions of the experimental list were constructed so that the same sentence did not always appear at the beginning or at the end of the experiment. The constraints on experimental list construction were that no target could be repeated over two successive trials, and that there were never more than five sentences in a row that contained the target.

Procedure. Each subject was tested individually in a quiet room. Printed instructions explained that they were to listen to a list of sentences. Before each sentence, a letter would be displayed in the centre of the screen: this was the target to be detected. Upon seeing the target, they were to think about the corresponding SOUND. Several examples with place names were given. For example, for the generalised phoneme-monitoring task: P as in

[3] The carrier nouns were monosyllabic ($n = 11$), bisyllabic ($n = 6$) or trisyllabic ($n = 5$).

[4] The carrier adjective was bisyllabic ($n = 4$), trisyllabic ($n = 14$) or quadrisyllabic ($n = 4$), and the preceding noun was always monosyllabic.

Paris, ePinal, gaP; K as in Canada, toKyo, afriQue; and for the word-initial phoneme-monitoring task: P as in Paris, Prusse, Pakistan; K as in Canada, Québec, Kyoto. Then, if the following sentence contained the sound that was represented by the visually presented letter, they were to press the response button AS FAST AS POSSIBLE. In the word-initial task, the subjects were warned that they were to press the button only if the target began a word, as in the examples given. If the sentence did not contain the sound, they were to do nothing and wait for the next target. It was emphasised that the target sound could appear anywhere in the sentence, and that if it appeared twice, they had to respond to the first one only. Speed and accuracy were emphasised.

The subjects were seated in front of a computer, wearing headphones, their preferred hand resting on the morse key that served to record reaction times. The auditory stimuli were stored at a sampling rate of 16 kHz and were presented directly through an OROS AU22 16-bit D/A board at 64 kHz (four times oversampled), followed by low-pass filtering at 20 kHz. A trial began with the visual presentation of a letter representing the target phoneme in the centre of the screen for 1 sec. The screen was left blank for another 1 sec. One sentence was presented. One and a half seconds after the end of the auditory presentation, the trial ended and a new trial began immediately. Pressure on the morse response key was recorded by the computer from the onset of the stimulus to the end of the trial. Response times were measured from the onset of all target phonemes.

Before the experiment began, the subjects performed a 10-item training trial. During this trial, the computer provided on-line feedback as to both correctness and reaction times of the responses.

Subjects. Altogether, 32 students from various Parisian universities, aged 20–26 years, were tested in each of the two variants of the experimental task (subjects were randomly assigned to one or the other version of the experimental task). Thirteen additional subjects were tested, but their data were discarded because their error rate was over 12%, taking into account misses and false alarms but not outliers (10 subjects in the generalised phoneme detection task, 3 in the initial detection task). None of the subjects had any known hearing deficit. All were native speakers of standard French.

Results

After discarding reaction times that were negative or greater than 3 sec, reaction times above or below two standard deviations from the mean were discarded and replaced by the mean for each condition. The misses and outliers represented 11.2% of the data. The false alarm rate was 8.1% in the generalised phoneme detection task; for the initial detection task, it was 1.2% for sentences without a target and 8.8% for sentences with a word-medial target. The results are displayed in Fig. 2.

Two analyses of variance were conducted on the reaction time and error data, one with subjects and one with items as the random factor. The ANOVA by subjects included two between-subjects factor—experimental task with two modalities (generalised and initial) and list with four modalities (counterbalancing factor)—and the within-subject factor target position (on noun *vs* adjective). The ANOVA by items included the between-items factor target position and the within-item factor experimental task. The reaction time analysis revealed a significant main effect of experimental task [$F_1(1,56) = 17.0$, $P<0.01$, MSe = 210492; $F_2(1,42) = 122$, $P<0.01$, MSe = 144711; min $F'(1,71) = 14.9$, $P<0.01$] and a main effect of target position [$F_1(1,56) = 472$, $P<0.01$, MSe = 715305; $F_2(1,42) = 121$, $P<0.01$, MSe = 491770; min $F'(1,63) = 96.3$, $P<0.01$]; in addition, there was a significant interaction between these two factors [$F_1(1,56) = 49.0$, $P<0.01$, MSe = 74035; $F_2(1,42) = 43$, $P<0.01$, MSe = 50901; min $F'(1,94) = 22.9$, $P<0.01$]. This interaction stems from the fact that there is no task effect for targets on nouns [33 msec, $t(56) = 1.63$, $P = 0.1$], while there is a significant task effect for targets on adjectives [130 msec, $t(56) = 6.0$, $P<0.01$]. The counterbalancing list factor had no main effect and did not interact with any other factor.

The same analyses on the error data (including misses and outliers) revealed the same trends, but less significantly so: main effect of experimental task [$F_1(1,56) = 6.6$, $P<0.02$, MSe = 258; $F_2(1,42) = 4.0$, $P<0.05$, MSe = 178; min $F'(1,85) = 2.49$, $P = 0.1$]; main effect of target position [$F_1(1,56) = 14.4$, $P<0.01$, MSe = 404; $F_2(1,42) = 1.3$, $P>0.1$, MSe = 277; min $F'(1,50) = 1.2$, $P>0.1$]; interaction between experimental task and target position [$F_1(1,56) = 9.2$, $P<0.01$, MSe = 258; $F_2(1,42) = 4.0$, $P<0.05$, MSe = 178; min $F'(1,76) = 2.9$, $P>0.01$].

Discussion

There was a different pattern of results for the two conditions. When the target was situated on the noun, just after the determiner, subjects were as fast in both experimental tasks: it did not take them longer to check that the target phoneme was word-initial than just to find the target phoneme. In contrast, when the target was situated on the adjective, just after the noun, subjects were slower to perform the initial phoneme detection task than the generalised one: they had to do some additional processing to verify that the target phoneme was word-initial.

These different patterns allow us to eliminate instantly some interpretations of the comparison between the tasks. If subjects had

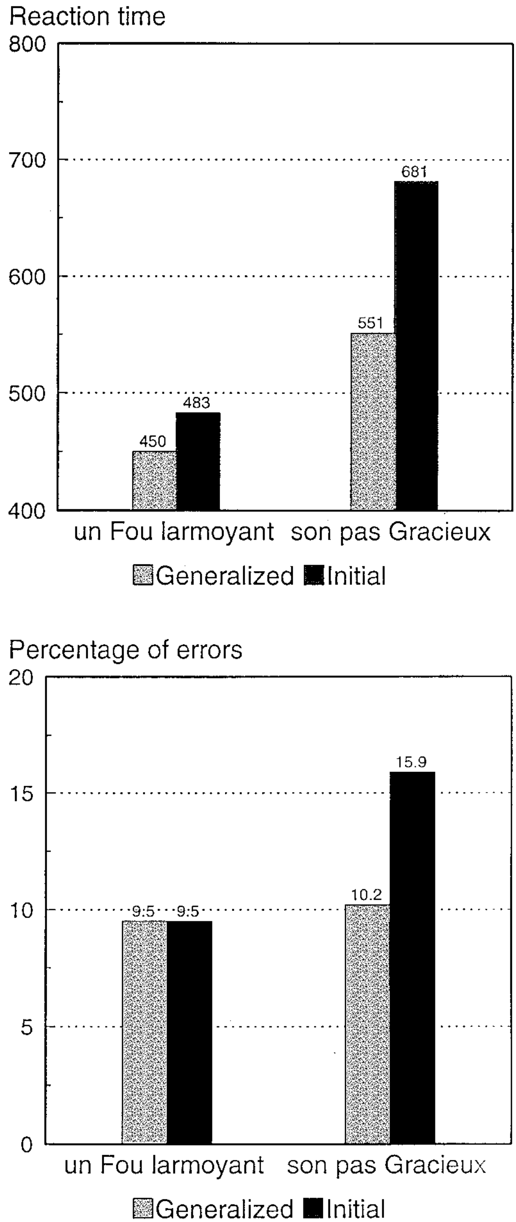

FIG. 2. Reaction time to targets situated on noun or on adjective in two experimental tasks: Initial phoneme detection and generalised phoneme detection.

systematically been slower for the word-initial phoneme detection task, we could not have been certain that a phonemic representation was available before word boundaries were known or even before lexical access was completed: it may just be that subjects were aware of the greater complexity of the task (the instructions stated “beware, you should answer only if the target phoneme is at the beginning of a word”). Therefore, they may have slowed down to double-check their responses. In contrast, if subjects had systematically been as fast and accurate for both tasks, we could not have concluded that word boundaries were available before lexical access was completed; rather, it may be that the representation that allows subjects to perform the phoneme detection task is available only after lexical access is completed. Therefore, the experiment would have informed us about the phoneme detection task in sentences (namely, that it can be performed only after lexical items are identified) but not about on-line lexical access from continuous speech.

However, the results as they stand allow us to make a firm conclusion about on-line processing: Given the same task requirements, some word boundaries are more readily available than others. Let us inspect closely the differences between the two conditions to detect what factors are responsible for the different patterns of results. In one case, the critical boundary is between a function word and a content word (determiner and noun), whereas in the other it is between two content words (noun and adjective). This is, in fact, the only difference that can account for the results: in both cases, the word preceding the critical boundary was monosyllabic, and in both cases the context was unambiguous, so that hearing the first syllable of the target word meant that it must be the first syllable of a new word. Thus, in “un Fou larmoyant”, there is no French word starting with the string of phonemes /ẽfu/; and in the case of “son pas gracieux”, there is no French word starting with the string of phonemes /pagra/ (and this was true of all sentences).

Our results allow us to draw the conclusion that determiners in noun phrases are accessed very quickly and that, as an immediate consequence, subjects know where the next content word begins.[5] In contrast, it seems to take some time to figure out that a monosyllabic noun has been heard when

[5] Given that we studied these effects with the phoneme detection task, we can also draw a conclusion about the representation used to perform this task (at least about one of them, in case there are several). Subjects not only have fast access to function words, allowing them to start lexical access with content words from the next syllable, but also the presence of a boundary has to be represented at the level used to respond to phonemes. Or, function words could be represented specially in that same representation (for instance, using a different code for each function word). This last hypothesis would be consistent with the fact that subjects seem to be deaf to phonemes in function words (see Wauquier-Gravelines, 1996), as pointed out to us by John Morton.

it is followed by an adjective (even though there is no ambiguity in the string of phonemes). The most plausible interpretation for this is that lexical access to this monosyllabic word is delayed: this implies that the boundary between the noun and the adjective is not marked pre-lexically. We know, for instance, that there is no phonotactic marking in these cases: the "pagra" sequence *could* be a word-initial sequence, even though it happens not to be. The results imply that there is no prosodic marking either (or if there is, subjects do not rely on it to make their decision). In fact, the whole NPs form phonological phrases in our sentences.[6] Lexical access to the monosyllabic noun, therefore, has to wait until some processes involving the knowledge of words, such as the competition between words implemented in models such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994), can establish the correct parsing into words for the string of phonemes following the function word.

The present results therefore demonstrate a special role for function words (at least for articles) during on-line lexical access from continuous speech in adult listeners. There is an extensive and controversial literature on the role of function words relative to content words (see, e.g. Bradley, 1978; Matthei & Kean, 1989; Petocz & Oliphant, 1988). However, few studies have investigated the processing of function words in sentence context in the auditory modality (as opposed to reading). For those studies that have done so, the trend seems to be in the direction of a special processing of function words, congruent with the present results (Cutler, 1993; Herron & Bates, in press; Swinney, Zurif, & Cutler, 1980). The present study probably represents the most direct evidence yet that function and content words already differ during lexical access, not only in post-access processes.

Several mechanisms are consistent with this special role for function words in on-line lexical access in adults. One of these is the function-word-stripping hypothesis (Fig. 1). On this view, both prosodic boundaries and function words play a role in on-line lexical access. As soon as they encounter a prosodic boundary, subjects compare the first few syllables to their list of function words; in case one matches, they identify it and start a lexical search from the next syllable. This procedure guarantees very fast access to the beginning of content words following a function word (itself following a prosodic boundary). However, this is not the only mechanism to account for the present results. For instance, subjects may access a function word whenever they encounter a syllable homophonous to that function word, and start a lexical search from the next syllable. This particular

[6] Starting from Nespor and Vogel's (1986) definition, the adjective restructures to form only one phonological phrase with the noun, and the whole NPs were never more than six syllables long.

proposal has the disadvantage of predicting that subjects should make many over-segmentation errors; a rough estimation for French shows that about a third of all content words have at least one of their syllables homophonous to a function word. To counter this, one could assume that function words are somehow phonetically marked (an assumption, however, that seems rather implausible for French). In this view, subjects would be able to identify real function words just by listening to them and no content word syllable would ever be truly homophonous to a function word. An alternative solution may be that subjects perform an on-line syntactic analysis that allows them to expect function words. Each of these proposals makes different predictions that can be tested experimentally.[7]

To summarise, we have shown that articles play a special role in on-line lexical access in adult speakers of a language. Several mechanisms can accommodate this special role. However, the one we favour, that prosodic boundary information and function words are used simultaneously in the process of lexical access, has the advantage of also offering an account of acquisition. Only further experiments will allow us to distinguish between alternative interpretations.

In conclusion, we would like to emphasise three points. The first is the richness of the phonological bootstrapping approach. We have shown here that the speech signal offers many cues to a language's structure, and that babies seem to be well equipped to process these cues. The second point concerns the advantages of considering more than one acquisition problem at a time: this approach may help us escape bootstrapping circularities (i.e. postulating knowledge in one domain to solve the acquisition problem in another and vice versa). In addition, it allows us to gather converging evidence from more than one domain of knowledge. For instance, phonological phrases and function words seem to be useful for lexical segmentation and acquisition, and for syntactic processing and acquisition. If the use of a given representation or process can be demonstrated in one condition, then all hypotheses relying on the same feature gain credibility. The last point we wish to emphasise is the richness of the research paradigm that considers simultaneously language acquisition by babies and speech processing by adults. On the one hand, such an approach enables us to constrain adults' models of speech processing so that there should exist a procedure by which the system can been acquired on the basis of an innate architecture and some linguistic input. On the other, it allows us to gather data on any given hypothesis from two populations of subjects—adults and

[7] For instance, the prosodic hypothesis predicts that subjects will be misled when encountering a syllable homophonous to a function word at the beginning of a prosodic unit (such as in the French verb "décider" where "des" is the plural definite article). The on-line syntactic analysis hypothesis can be tested by constructing sentences with different probabilities for given phrase types (a parameter that was not manipulated in the present study).

babies. On-line tasks can be used only with adults (at least to our present knowledge) and are crucial to settle issues of processing. And again, converging evidence with a variety of tasks can be obtained.

REFERENCES

Bailly, G. (1989). Integration of rhythmic and syntactic constraints in a model of generation of French prosody. *Speech Communication*, *8*, 137–146.

Bloom, L. (1970). *Language development: Form and function in emerging grammars.* Cambridge: MIT Press.

Bradley, D.C. (1978). *Computational distinctions of vocabulary type*. Unpublished PhD thesis, MIT, Boston, MA.

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

Brown, R. (1973). *A first language*. Cambridge: Harvard University Press.

Cairns, P., Shillcock, Chater, & Levy, (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology*, *33*, 111–153.

Childers, J.B., & Echols, C.H. (1996). Infants' use of rhythmic cues in word-level segmentation. In A. Stringfellow, D. Cahana-Amitay, E. Hughes, & A. Zukowski (Eds), *Proceedings of the 20th Boston University Conference on Language Development*, pp. 167–176. Somerville, MA: Cascadilla Press.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N.A., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Christophe, A. (1993). *Rôle de la prosodie dans la segmentation en mots*. EHESS: Paris.

Christophe, A., & Dupoux, E. (1996). Bootstrapping lexical acquisition: The role of prosodic structure. *Linguistic Review*, *13*, 383–412.

Christophe, A., Dupoux, E., & Mehler, J. (1993). How do infants extract words from the speech stream? A discussion of the bootstrapping problem for lexical acquisition. In E.V. Clark (Ed.), *Proceedings of the 24th Annual Child Language Research Forum*, pp. 209–224. Stanford, CA: CSLI.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, *95*, 1570–1580.

Cutler, A. (1993). Phonological cues to open- and closed-class words in the processing of spoken sentences. In special issue: Sentence processing III. *Journal of Psycholinguistic Research*, *22*, 109–131.

Cutler, A. (1996). Prosody and the word boundary problem. In J.L. Morgan & K. Demuth (Eds), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, pp. 87–99. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*, 218–236.

Cutler, A., & Carter, D.M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, *2*, 133–142.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 113–121.

Delais-Roussarie, E. (1995). *Pour une approche parallèle de la structure prosodique: Etude de l'organisation prosodique et rythmique de la phrase française*. Unpublished PhD thesis, Université de Toulouse-Le Mirail, Toulouse.

de Pijper, J.R., & Sanderman, A.A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *Journal of the Acoustical Society of America*, *96*, 2037–2047.

Dirksen, A. (1992). Accenting and deaccenting: A declarative approach. In *Proceedings of the 15th International Conference on Computational Linguistics*, Vol. 3, pp. 865–869. Nantes:

Friederici, A.D., & Wessels, J.M.I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech-perception. *Perception and Psychophysics*, *54*, 287–295.

Gerken, L. (1994). Young children's representation of prosodic phonology: Evidence from English-speakers' weak syllable productions. *Journal of Memory and Language*, *33*, 19–38.

Gerken, L. (1996). Phonological and distributional information in syntax acquisition. In J.L. Morgan & K. Demuth (Eds), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, pp. 411–425. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Gerken, L., & McIntosh, B. (1993). Interplay of function morphemes and prosody in early language. *Developmental Psychology*, *29*, 448–457.

Gerken, L., Landau, B., & Remez, R.E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology*, *26*, 204–216.

Gerken, L., Jusczyk, P.W., & Mandel, D.R. (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition*, *51*, 237–265.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55.

Gleitman, L., & Wanner, E. (1982). The state of the state of the art. In E. Wanner & L. Gleitman (Eds), *Language acquisition: The state of the art*, pp. 3–48. Cambridge: Cambridge University Press.

Goodsitt, J.V., Morgan, J.L., & Kuhl, P.K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language*, *20*, 229–252.

Gow, D.W., & Gordon, P.C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 344–359.

Hayes, B., & Lahiri, A. (1991). Bengali intonational phonology. *Natural Language and Linguistic Theory*, *9*, 47–96.

Herron, D., & Bates, E. (in press). Grammatical and acoustic factors in the recognition of open- and closed-class words. *Journal of Memory and Language*.

Hirsh-Pasek, K., & Golinkoff, R.M. (1996). *The origins of grammar: Evidence from early language comprehension*. Cambridge, MA: MIT Press.

Hirsh-Pasek, K., Nelson, D.G.K., Jusczyk, P.W., Cassidy, K.W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, *26*, 269–286.

Jusczyk, P.W., & Thompson, E.J. (1978). Perception of a phonetic contrast in multisyllabic utterances by two-month-old infants. *Perception and Psychophysics*, *23*, 105–109.

Jusczyk, P.W., Cutler, A., & Redanz, N.J. (1993a). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*, 675–687.

Jusczyk, P.W., Friederici, A.D., Wessels, J.M.I., Svenkerud, V.Y., & Jusczyk, A.M. (1993b). Infants' sensitivity to the sound pattern of native language words. *Journal of Memory and Language*, *32*, 402–420.

Klatt, D.H. (1979). Speech perception: A model of acoustic–phonetic analysis and lexical access. *Journal of Phonetics*, *7*, 279–312.

Kuhl, P. (1992). Innate predispositions and the effets of experience in speech perception: The native language magnet theory. In B.D. Boysson-Bardies, S.D. Schonen, P.W. Jusczyk, P. McNeilage, & J. Morton (Eds), *Developmental neurocognition: Speech and face processing in the first year of life*, pp. 259–274. Dordrecht: Kluwer Academic.

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

Marslen-Wilson, W.D., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes and features. *Psychological Review*, *101*, 653–675.

Matthei, E.H., & Kean, M.L. (1989). Postaccess processes in the open *vs* closed class distinction. *Brain and Language*, *36*, 163–180.

Mazuka, R. (1996). How can a grammatical parameter be set before the first word? In J.L. Morgan & K. Demuth (Eds), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, pp. 313–330. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

McClelland, J.L., & Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McQueen, J.M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 621–638.

Mehler, J., Dupoux, E., & Segui, J. (1990). Constraining models of lexical access: The onset of word recognition. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*, pp. 263–280. Cambridge, MA: Bradford Books.

Morgan, J.L. (1986). *From simple input to complex grammar*. Cambridge, MA: MIT Press.

Morgan, J.L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language*, *35*, 666–688.

Morgan, J.L., & Demuth, K. (1996a). Signal to syntax: An overview. In J.L. Morgan & K. Demuth (Eds), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, pp. 1–22. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

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

Morgan, J.L., Shi, R., & Allopenna, P. (1996). Perceptual bases of rudimentary grammatical categories: Toward a broader conceptualization of bootstrapping. In J.L. Morgan & K. Demuth (Eds), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, pp. 263–283. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*: Dordrecht: Foris.

Nespor, M., Guasti, M.T., & Christophe, A. (1996). Selecting word order: The Rhythmic Activation Principle. In U. Kleinhenz (Ed.), *Interfaces in phonology*, pp. 1–26. Berlin: Akademie Verlag.

Newsome, M.R., & Jusczyk, P.W. (1995). Do infants use stress as a cue in segmenting fluent speech? In C. MacLaughlin & S. McEwen (Eds), *Proceedings of the 19th Boston University Conference on Language Development*, Vol. 2, pp. 415–426. Boston, MA: Cascadilla Press.

Norris, D.G. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Norris, D., McQueen, J.M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 621–638.

Oviatt, S.L. (1980). The emerging ability to comprehend language: An experimental approach. *Child Development*, *51*, 97–106.

Pasdeloup, V. (1990). *Modèle de règles rythmiques du français appliqué à la synthèse de la parole*. Unpublished PhD thesis, Université d'Aix-en-Provence, Aix-Marseille.

Petocz, A., & Oliphant, G. (1988). Closed-class words as first syllables do interfere with lexical decisions for nonwords: Implications for theories of agrammatism. *Brain and Language*, *34*, 127–146.

Pinker, S. (1994). *The Language Instinct*. London: Penguin Books.

Radford, A. (1990). *Syntactic theory and the acquisition of English syntax*. Oxford: Blackwell.

Roberts, K. (1983). Comprehension and production of word order in stage I. *Child Development*, *54*, 443–449.

Sansavini, A. (1995). Newborns perceive rhythmical units of speech: The role of stress and vowels. Paper presented at the *Second SISSA Workshop on Processing Consequences of Contrastive Phonologies*, Trieste, Italy, May.

Shafer, V.L., Gerken, L.A., Shucard, J.L., & Shucard, D.W. (1992). "The" and the brain: An electrophysiological study of infants' sensitivity of English function morphemes.

Swinney, D.A., Zurif, E.B., & Cutler, A. (1980). Effects of sentential stress and word class upon comprehension in Broca's aphasics. *Brain and Language*, *10*, 132–144.

Tabossi, P., Burani, C., & Scott, D. (1995). Word identification in fluent speech. *Journal of Memory and Language*, *34*, 440–467.

Thomas, D.G., Campos, J.J., Shucard, D.W., Ramsey, D.S., & Shucard, J. (1981). Semantic comprehension in infancy: A signal detection analysis. *Child Development*, *52*, 798–803.

Wauquier-Gravelines, S. (1996). *Organisation phonologique et traitement de la parole continue: Contraintes prosodiques et phonologiques de l'accès lexical*. Unpublished PhD thesis, Université Paris VII, Paris.

Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P.J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, *91*, 1707–1717.