JOURNAL OF MEMORY AND LANGUAGE **39,** 246–263 (1998)
ARTICLE NO. ML982576

PARSER: A Model for Word Segmentation

Pierre Perruchet and Annie Vinter

University of Bourgogne, Dijon, France

Saffran, Newport, and Aslin (1996b) showed that adults were able to segment into words an artificial language that included no pauses or other prosodic cues for word boundaries. We propose an account of their results that requires only limited computational abilities and memory capacity. In this account, parsing emerges as a natural consequence of the on-line attentional processing of the input, thanks to basic laws of memory and associative learning. Our account was implemented in a computer program, PARSER. Simulations revealed that PARSER extracted the words of the language well before exhausting the material presented to participants in the Saffran et al. experiments. In addition, PARSER was able to simulate the results obtained under attention-disturbing conditions (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997) and those collected from 8-month-old infants (Saffran, Aslin, and Newport, 1996a). Finally, the good performance of PARSER was not limited to the trisyllabic words used by Saffran et al., but also extended to a language composed of one- to five-syllable words. © 1998 Academic Press

Perceiving a word as a unit may appear trivial to literate people, because words are displayed in isolation in written language. However, language acquisition initially proceeds from auditory input, and linguistic utterances usually consist of sentences linking several words without clear physical boundaries. The question thus arises: How do infants become able to segment a continuous speech stream into words?

Recent psycholinguistic research has identified a number of potentially relevant factors. Analyses of the statistical structure of different languages have shown that a number of features are correlated with the presence of word boundaries and could therefore be used as cues to segment the speech signal into words. The prosodic characteristics of discourse, such as pitch and stress, provide such cues. In English, for instance, strong syllables tend to initiate words. Correspondingly, English-speaking adults appear to use prosodic cues such as strong and weak syllables to parse a continuous acoustic signal into words (Cutler & Butterfield, 1992). In addition, phonological information may be relevant. For instance, the pronunciation of phonemes may differ within and across word boundaries, a phenomenon known as allophonic variation. Each language is also characterized by a set of phonotactic regularities, which describe which phonemes and sequence of phonemes are likely to appear in different syllabic positions. Experimental studies have shown that infants are sensitive to both prosodic and phonological regularities of their native language, suggesting that these cues may be used during language acquisition (see Brent & Cartwright, 1996; Christiansen, Allen & Seidenberg, in press; Jusczyk, 1997; McDonald, 1997).

However, one may question whether exploitation of the statistical regularities of a language provides a definitive response to the question of how infants extract words. Indeed, hypothesizing that these cues guide segmentation only transposes the learning problem one step back. The question remains as to how infants abstract the statistical regularities that they seemingly exploit. One cannot claim that these regularities

This research was supported by the University of Bourgogne and the Centre National de la Recherche Scientifique (ESA CNRS 5022). We thank Axel Cleeremans and Ronald Peereman for discussing the ideas presented in this paper, as well as Arthur Samuel and two anonymous reviewers for their careful reviews and helpful suggestions on a previous draft.

Address correspondence and reprint requests to either author at LEAD/CNRS, Faculte des Sciences, 6 Bd Gabriel, 21000, Dijon, France. E-mail: perruche@u-bourgogne.fr, vinter@u-bourgogne.fr.

0749-596X/98 $25.00

are learned inductively from word exposure without falling into circular reasoning, with word knowledge being simultaneously the prerequisite and the consequence of knowledge of statistical regularities. Partial solutions have been proposed to avoid circularity, notably in the case of rhythmic cues, which have been shown to play a role in word segmentation for native speakers of stressed languages (e.g., Echols, Crowhurst, & Childers, 1997). If the alternation of strong and weak syllables provides a guide for segmentation, listeners must only infer whether the stress is word-initial, as in English, or falls on another syllable, as in some other languages. Arguably, this and other specifications could be learned when words are presented in isolation. Also, Brent and Cartwright (1997) suggested that infants learn the constraints existing for the beginning and the end of words from global utterances, assuming that the sequences permissible at word boundaries are the same as the ones that occur at utterance boundaries. However, basing word segmentation on such indirect mechanisms remains somewhat unsatisfactory. In addition to the difficulties inherent in their exploitation, prosodic and phonological cues in any case provide only probabilistic information.

The importance of prosodic and phonological cues in word discovery is further questioned by recent experimental studies showing that these cues are not necessary, at least for adults learning to segment an artificial language (e.g., Saffran, Newport, & Aslin, 1996b). Because these studies are the target of the following simulations, we will review them in some detail. Saffran et al. (1996b) used an artificial language composed of six trisyllabic words, such as *babupu* and *bupada.* The words were presented in random order and repeated for 21 min. They were read by a speech synthesizer in immediate succession, without pauses or any other prosodic cues. Thus participants heard a continuous series of syllables without any word boundary cues. Note that the lack of pauses and prosodic cues also prevented participants from abstracting other statistical regularities relevant for segmentation, insofar as this operation would require at least a few words to have been previously isolated. In the following phase, participants received a forced choice test in which they had to indicate which of two items sounded more like a word from the artificial language. One of the items was a word from the artificial language, whereas the other was a new combination of three syllables belonging to the language. Participants selected the correct words on 65 or 76% of the trials, depending on whether the alternative item included a permissible syllable pair. In both cases, participants performed significantly better than would be expected by chance.

These results suggest that people are able to learn the words composing a language without any prosodic cues. However, the participants in the study of Saffran et al. (1996b) were told before the training session began that the artificial language contained words, and they were asked to figure out where the words began and ended. The processes used in these conditions may be different from those involved in natural language acquisition. Two subsequent papers from the same laboratory (Saffran, Aslin, & Newport, 1996a; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997) partially responded to this objection. In Saffran et al. (1997), the primary task of participants was to create an illustration with a coloring program. They were informed that an audiotape that would be playing in the background might affect their artistic creativity and that the experiment was aimed at investigating this influence. They were not told that the tape consisted of a language nor that they would be tested later in any way. In the subsequent forced choice test, participants selected the words in 58.6% of the trials, a score significantly better than chance. In a second experiment, Saffran et al. (1997) replicated the same procedure, but participants colored and listened to the 21-min tape during two sessions instead of one. In these conditions, the mean score reached 73.1%.

All of the results described so far were collected from adults. Saffran et al. (1997) also studied a group of 6- and 7-year-old children. The overall performance of children was not significantly different from that of adults. However, a more direct indication of the relevance of these data with regard to infants acquiring their mother tongue was provided by Saffran et al. (1996a), who reported studies carried out with 8-month-old infants. In order to adapt the task to infants, there were only four trysyllabic words, and the duration of the familiarization phase was reduced to 2 min. The infants were subsequently tested with the familiarization-preference procedure of Jusczyk and Aslin (1995), in which infants controlled the exposure duration of the stimuli by their visual fixation on a light. The infants showed longer fixation (and hence listening) times for nonwords than for words, an effect the authors attributed to novelty preference. The mean difference between words and nonwords, although relatively small (approximately 850 ms over a total exposure time of about 8 s) was statistically significant, demonstrating that infants were sensitive to word structure after a brief exposure to an artificial language.

Overall, the studies conducted by Saffran and co-workers offer impressive support for the hypothesis that people are able to segment a continuous speech stream without any prosodic or phonological cues for word boundaries. That is not to say that these cues are not exploited for word segmentation. However, as pointed out earlier, exploiting the statistical regularities associated with the word-level organization of the language logically implies prior knowledge of at least some words, and hence could not be construed as the single or primitive solution to the word segmentation issue. The phenomenon described by Saffran and co-workers may allow formation of an initial knowledge base from which these regularities can be subsequently inferred and exploited. The mechanisms underlying the intriguing ability revealed in these studies remain to be determined.

Saffran et al. (1996b) pointed out that the only cues available to participants were the distributional statistics of subunits. More precisely, analysis of the structure of any language, whether natural or artificial, shows that the correlations between contiguous syllable pairs belonging to a word is stronger, on average, than the correlations between contiguous syllables which straddle word boundaries. The authors hypothesized that infants exploited this universal property. In their conception, infants inferred word boundaries from the discovery of troughs in the distribution of the transitional probabilities between syllables.

As the authors themselves noted, however, this task represents ‘‘a computational endeavor of considerable complexity’’ (Saffran et al., 1996b, p. 610). In the remainder of this paper, we argue that participants succeeded in the segmentation task without computing transitional probabilities. We argue that parsing linguistic input into meaningful units becomes fairly simple when, instead of thinking about the problem as a statistician, one considers how it can be solved by the human processing system. Starting from the view that parsing an unknown language represents an instance of implicit learning, we use an approach that has evolved primarily in the implicit learning area (e.g., Perruchet & Vinter, 1998; Perruchet, Vinter, & Gallego, 1997; Vinter & Perruchet, 1997; see also Perruchet & Pacteau, 1990, for a first draft of our account). We first outline the principles underlying our approach to parsing and the lineaments of PARSER, a computer model implementing these principles. Then we present four simulations. The first three simulations concern the data reported by Saffran et al. (1996a, 1996b, 1997), and the fourth tests the effect of some procedural variations on the efficiency of PARSER.

A NEW ACCOUNT OF PARSING

Let us start with a common observation. When people are faced with material composed of a succession of elements, each of them matching some of the people's processing primitives, they segment this material into small and disjunctive parts comprising a few primitives. As adults, we have direct evidence of the phenomenon. For instance, when asked to read nonsense consonant strings, we read the material not on a regular rhythmic, letter-by-letter basis, but rather by chunking a few letters together. In a more experimental vein, when adults are asked to write down this kind of material, they frequently reproduce the strings as separate groups of two, three, or four letters

(Servan-Schreiber & Anderson, 1990). The same phenomenon may occur when a listener is faced with an unknown spoken language, with the syllables forming the subjective processing primitives instead of the letters.[1] Chunking, we contend, is a ubiquitous phenomenon, due to the intrinsic constraints of attentional processing, with each chunk corresponding to one attentional focus (for a theoretical justification of this principle, see Perruchet & Gallego, 1997).

It could be argued that taking this phenomenon as a starting point is like taking what we intend to explain as a premise. The argument is correct in the sense that we do not attempt to account for initial chunking other than through its dependence on the properties of attentional processing. However, what remains to be explained is how the chunks turn out to match the words of the language. Indeed, initial segmentation is assumed to depend on a large variety of factors. Some factors are linked to the participants, such as their prior experience with another language and their state of attention and vigilance. Other factors are linked to the situation, such as the signal/noise ratio, the time parameters of the speech signal, and the relative perceptual saliency of the components of the signal. The mixture of these factors is very likely to make the initial chunks different in length and content from the words of the language.

Our proposal is that correct parsing emerges as a direct consequence of the organization of the cognitive system. For convenience, this organization may be characterized by the interplay of two interrelated principles. The first principle stipulates that perception shapes internal representations. With regard to our concern, this means that the primitives that are perceived within one attentional focus as a consequence of their experienced spatial or temporal proximity become the constituents of one new representational unit. The future of this representational unit, we posit, depends on ubiquitous laws of associative learning and memory. If the association between the primitives which form a percept is not repeated, the internal representation created by this percept rapidly vanishes, as a consequence of both natural decay and interference with the processing of similar material. However, if the same percept reoccurs, the internal representation is progressively strengthened. The second principle is that internal representations guide perception. Perception involves an active coding of the incoming information constrained by the perceiver's knowledge. Because this knowledge changes with increased experience in a domain, perception, and notably the composition and the size of the perceived chunks, itself evolves. (This view, which contrasts with the claim that perception is driven by a fixed repertoire of primitive features, has been cogently documented by Schyns, Goldstone, and Thibaut, 1998; see also Goldstone, Schyns, & Medin, 1997). The resulting picture is that perception builds the internal representations which, in turn, guide further perception.

The relevance of these general principles becomes clear when they are considered jointly with a property inherent to any language. If the speech signal is segmented into small parts on a random basis, these parts have more chance of being repeated later if they match a word, or a part of a word, than if they straddle word boundaries.

How does the system work? We saw that perception naturally segments the material into disjunctive parts. This phenomenon provides the system with a sample of potential units, some of them relevant to the structure of the language and others, presumably most, irrelevant. According to the first principle described above, an internal representation which matches a percept is reinforced in the system if the percept occurs repeatedly. Given the above-mentioned property of language, this means that lasting internal representations are more likely to match a word or part of a word than a between-word segment. The relevant units, in

[1] The choice of syllables as a natural unit of speech processing is controversial. However, considerable evidence exists to support this postulate (e.g., Finney, Protopapas, & Eimas, 1996; see review in Eimas, 1997; Jusczyk, 1997). Of special relevance here is that infants are able to chunk speech samples into syllables and to represent these syllables for at least short time intervals (e.g., Jusczyk, Kennedy, & Jusczyk, 1995). The choice of syllables was also motivated by our wish to achieve the best possible fit with the Saffran et al. studies. However, this choice is not essential for the model, insofar as the same line of reasoning could apply with phonemic-level representations as input, for instance.

this sketch, emerge through a natural selection process, because forgetting and interference lead the human processing system to select the repeated parts among all of the parts generated by the initial, presumably mostly irrelevant, chunking of the material. The second principle ensures the convergence of the process toward an optimal parsing solution. The fact that perception is guided allows the system to build representations of words whose components could hardly be perceived in one attentional focus if perception were driven only by the initial primitives of the language. Also, once internal representations providing an appropriate coding of the input have been built, an endless generation of new irrelevant percepts is avoided.

The components of our account, when considered individually, are far from new. This observation obviously holds for the property of the language on which this account relies. It also holds for the psychological principles involved. We have recourse to no new and specialized learning devices. For instance, the unitization of elements due to their processing within the same attentional focus is one of the basic tenets of associative learning (e.g., Mackintosh, 1975). Likewise, the laws of forgetting and the effects of repetition are ubiquitous phenomena in memory. Moreover, the mutual dependence of perception and internal representations, which is the cornerstone of our account, is in line with a developmental principle initially described by Piaget's concepts of assimilation and accommodation (e.g., Piaget, 1985). Most current theories of development, although they use different terminology, also rely on the constructive interplay between assimilation-like and accommodation-like processes (e.g. Case, 1993; Fischer & Granott, 1995; Karmiloff-Smith, 1992). Our objective is to show that the segmentation into words of a continuous speech stream can be explained within a general view of human information processing.

To summarize, we propose that parsing results from the interaction between one property of the language—essentially that the probability of repeatedly selecting the same group of syllables by chance is higher if these syllables form intra-word rather than between-words components—and the properties of the processing systems—essentially that repeated percepts shape new lasting representations, which are in turn able to guide perception. To our knowledge, this fairly simple solution to the parsing issue has not yet been evaluated.

PARSER: A COMPUTER MODEL

PARSER is intended to show that our account of how words are extracted from continuous speech works as we anticipate when implemented in a computer simulation.

The Model

PARSER is centered around a single vector, called percept shaper (PS). PS is composed of the internal representations of the displayed material and may be thought of as a memory store or a mental lexicon. A weight, which reflects the person's familiarity with the item, is assigned to each element in PS. At the start of the familiarization session, PS contains only the primitives needed for processing the material (i.e., a few syllables). At the end, it should contain, in addition, all the words and legal clauses (i.e., units combining a few words) of the language. During the shaping process, PS may contain a mixture of words and legal clauses, part-words (e.g., two syllables out of a three-syllable word), and nonwords (e.g., the last syllable of one word with the first syllable of another word).

The way the words are built in PS during training is described in the flowchart in Fig. 1. Let us consider how the flowchart works for Simulation 1 of Study 1 reported below, which concerns the Saffran et al. (1996b) studies. Six trisyllabic words, *babupu, bupada, dutaba, patubi, pidabu,* and *tutibu,* were repeated in random order. For Simulation 1, the sequence began with: *tutibudutabatutibupatu* . . . The string was first segmented into small and disjunctive parts. In PARSER, the multiple determinants of this initial parsing were simulated by a random generator, which selected the size of the next percept within a range of 1 to 3 units (Fig. 1, step a). For Simulation 1, the random generator provided 2, 3, 2, 3, and 1 in the first five trials. In consequence, the first percepts were *tuti, buduta, batu, tibupa,* and *tu.* Because none of the first four percepts was present in PS (step b),

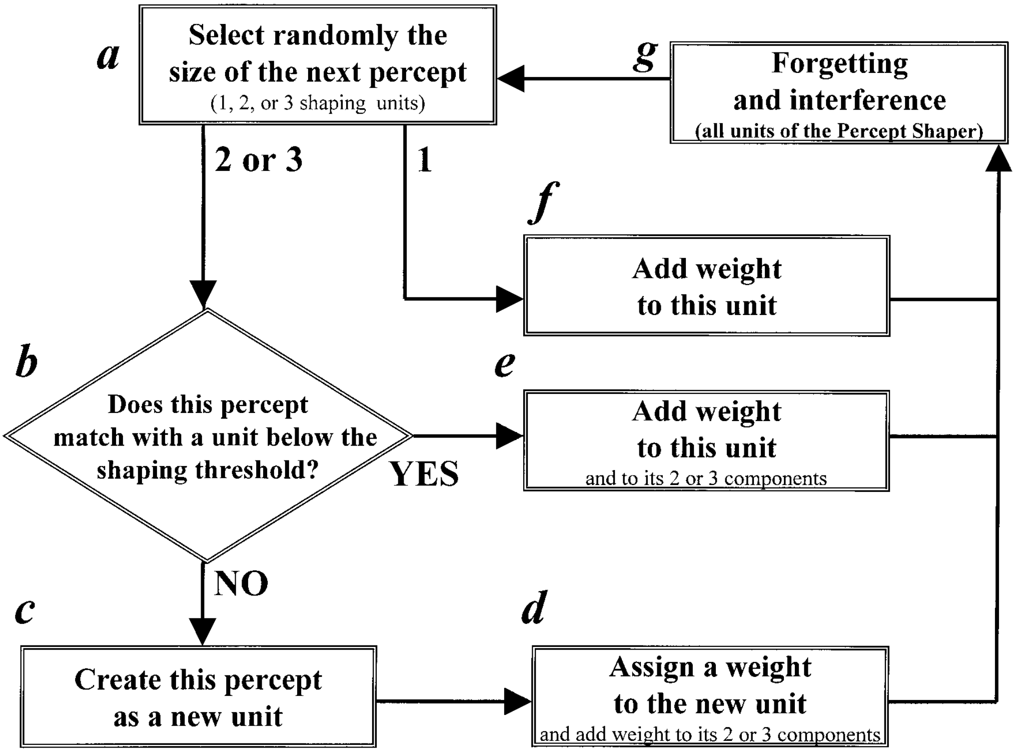
a
Select randomly the size of the next percept
(1, 2, or 3 shaping units)
2 or 3
1
g
Forgetting and interference
(all units of the Percept Shaper)
f
Add weight to this unit
b
Does this percept match with a unit below the shaping threshold?
YES
e
Add weight to this unit
and to its 2 or 3 components
NO
c
Create this percept as a new unit
d
Assign a weight to the new unit
and add weight to its 2 or 3 components

FIG. 1. Operations performed by PARSER at each time step.

they were created as new units (step c) and assigned a weight (step d). Also, the weights of the components, *tu, ti, bu,* and so on, were incremented. The fifth percept, *tu,* matched a primitive and hence was already represented in PS. Its weight was also incremented (step f).

At each time step (a time step is defined by the processing of a new percept, that is, by one cycle in the flowchart in Fig. 1), the units forming PS are affected by forgetting and retroactive interference (Fig. 1, step g). Forgetting is simulated by decreasing all the units by a fixed value. Interference is simulated by decreasing the weights of the units in which any of the syllables involved in the currently processed unit are embedded. In the case described here, interference occurred for the first time while *batu* was perceived. Indeed, *tu* was already present in two old units in PS, *tu* and *tuti.* In consequence, the weights of *tu* and *tuti* were decremented when *batu* was perceived (in addition to the decrement due to forgetting).

In this early phase, perception was driven by the initial primitives of the system, namely the syllables. However, the psychological principles implemented by the model stipulate that a representation created during learning may become able to guide perception, as the initial primitives were. The condition for an element of PS to shape perception is that its weight is at least equal to a threshold value. In contrast, when the frequency of perceiving a given element is not high enough to counteract the effects of forgetting and interference, this element is removed from PS when its weight becomes null.

In the reported simulations, the starting weight given to any created unit, was set to 1. The primitives that formed PS before training were also assigned a weight of 1. The increment received by an old unit in PS when this unit serves to shape perception, was set to 0.5. The decrements due to forgetting and interference were set to 0.05 and 0.005, respectively (except in Study 3). The last parameter, namely the threshold above which a

TABLE 1

Changes in Constituents of PS during One Time Step

	Initial		Processing components (add weight, interference)				Final
	weights	Creation	ba	bupa	da	Forgetting	weights
bu	3.50			−.005		− .05	3.445
pa	3.10			−.005		− .05	3.045
ba	2.55		+.5			− .05	3
du	2.50					− .05	2.45
ta	2.50					− .05	2.45
tu	2.45					− .05	2.40
da	2.40				+.5	− .05	2.85
bi	1.70					− .05	1.65
bupa	1.58			+ .5		− .05	2.03
pu	1					− .05	0.95
puduta	1					− .05	0.95
babupada	—	1					1

Note. In this example, the changes are due to the perception of *babupada,* when *ba, bupa,* and *da* are existing units of the system.

unit is able to shape perception, was set to 1. Although the absolute values of these parameters were arbitrary, their relations, which were the only relevant aspects for the model, exhibit at least rough behavioral likelihood. For instance, the parameters were set in such a way that a unit, when just created in PS, is only able to shape the immediately subsequent percept. Indeed, its initial weight (1) is quickly decreased due to forgetting (0.05), and it therefore no longer meets the threshold value for shaping perception (1). If this unit is not perceived again within the next 20 (1/0.05) percepts, its weight becomes null and it will be eliminated from PS (note that interference may speed the process). A unit that has gained a weight of 3, for instance, which means that it has been perceived from 4 to 10 times or more (each new percept involving this unit is accompanied by a gain of 0.5, but the effects of forgetting and interference are unpredictable, because they depend on the number and the nature of the intervening percepts), will lose its ability to shape new percept within 40 time steps and will disappear from PS within 20 further time steps (again when interference is ignored).

Let us return now to the performances observed in Simulation 1 to examine how the system works quantitatively. Table 1 shows the content of PS and the weight of each unit (columns 1 and 2) after the processing of 44 syllables (PS also included two primitives with their initial weights, and units with a weight lower than 1, which are not reproduced here). The continuation of the sequence was *babupadapida...* The random generator (Fig. 1, step a) determining the number of units embedded in the next percept provided a value of 3. Perception was shaped by the three units *ba, bupa,* and *da.*[2] Because *babupada* did not match a unit the

[2] Note that *bupa* guided perception at the expense of *bu.* This is because PARSER selects the longest unit when several candidates are possible. Other options would have been possible, such as selecting the most highly weighted unit or drawing randomly among the candidates. We ran pilot simulations implementing those options. Overall, they tended to perform worse than the model described here. Although selecting the longest units is consonant with COHORT (Marslen-Wilson & Welsh, 1978), our choice entails no adhesion to this specific model of spoken word recognition. Indeed, models of word recognition such as COHORT, TRACE (McClelland & Elman, 1986), or SHORTLIST (Norris, 1994) address the question of how speech is segmented by adult listeners who have an internal lexicon. The issue addressed by PARSER is quite different, insofar as it concerns the *creation* of the lexicon.

TABLE 2

Composition of PS Corresponding To What Is Referred To Hereafter as the *Loose Criterion*[a] for Simulation 1 of Study 1

Unit	Weight
pidabu	21.57
dutaba	21.10
babupu	17.39
bupada	16.39
tutibu	14.60
patubi	13.75
bu	13.68
tu	5.74
tuti	4.87
pa	1.88
tutibupatubi	1

[a] All the words of the language are in the highest part of PS, but PS also contains illegal units.

weight of which was below the shaping threshold (step b), this percept was created as a unit in PS (step c) and assigned a weight of 1 (step d; see Table 1, column 3). In addition, the components forming *babupada* received an additional weight of 0.5 (Table 1, columns 4, 5, and 6). Note that *bupa* was incremented, in keeping with the fact that it was an autonomous component of the percept *babupada,* but not its primitive parts *bu* and *pa,* that did not share this status. On the contrary, *bupa* interfered with *bu* and *pa* and, in consequence, the weights of these units were decremented by .005. Finally, all the units in PS were decremented by 0.05 to simulate forgetting (Fig. 1, step g; see Table 1, column 7). The rightmost column of Table 1 displays the state of PS after *babupada* has been processed.

The Performance Criteria

In order to introduce the criteria of learning used in the simulations, consider now the content of PS at later stages of learning. In Tables 2 and 3, the units unable to shape perception (i.e., with weights lower than 1) are not reported. The primitives whose weight was unchanged from their initial weight (i.e., 1) are also omitted. Reporting these primitives should be uninformative, because, by construction, the primitives were maintained in PS in order to allow the processing of new words that could be introduced during the course of training (this possibility will not be exploited in the reported simulations).

Table 2 shows the content of PS after processing 1347 syllables. The six words of the language were now in PS and were assigned the highest weights. However, PS also included a few items that were not words or legal clauses, such as *bu.* In the following simulations, this state of PS defines what will be referred to as the *loose criterion.*

Table 3 presents the state of PS at a still later stage of learning, after processing 2730 syllables. The six elements with the strongest weights were the six words forming the language, as before, but in addition all the remaining items were now legal clauses (all the primitives are assigned a weight of 1). Hereafter, this state of PS defines the *strict criterion.* With continued practice, the state of PS does not change further in any significant way. Because the perception is entirely shaped by internal representations, the weight of individual words continues to increase. New legal clauses are also continuously created, then quickly forgotten. The clauses are short-lived because all the possible successions of words were allowed,

TABLE 3

Composition of PS Corresponding To What Is Referred To Hereafter as the *Strict Criterion*[a] for Simulation 1 of Study 1

Unit	Weight
dutaba	48.53
bupada	42.78
babupu	39.09
pidabu	37.58
tutibu	37.24
patubi	36.65
tutibupatubi	1.16
bupadapidabu	1

[a] All the words of the language are in the highest part of PS, and the residual units are legal clauses.

thus making clauses uninformative about the structure of the language. If the generated sequence had followed syntactic rules preventing some possible successions, PARSER may have tended to select and strengthen syntactically correct utterances.

QUANTITATIVE STUDIES

Study 1

Let us first present the results of the study used in the above illustration. The situation under scrutiny was designed by Saffran et al. (1996b), Exp. 1). Recall that six trisyllabic words (*babupu, bupada, dutaba, patubi, pidabu,* and *tutibu*) were read in random order by a speech synthesizer. The participants heard a continuous series of 4536 syllables without any word boundary cues. The series of syllables presented to PARSER was obtained by concatenating the trisyllabic words in random order, the only restriction being that the same word was not repeated twice in immediate succession. The other random parameter of the model was the number of primitives forming a given unit (from 1 to 3). Because of the stochastic nature of the program, 100 simulations were run for this and all the studies reported in this paper, with the random parameters differing for each simulation. The results were processed as the results of 100 individual participants. The fixed parameters were set as mentioned above.

The loose criterion was reached after processing a mean of 2064 syllables, with a range of 735 to 4479 syllables. As mentioned above, the loose criterion is defined by the fact that the six words are formed in PS and have the highest weights. However, PS also included other items at this stage, and notably part-words. The second, more stringent criterion, is defined by the fact that the only residual items in PS are legal clauses, that is, strings composed of a few words. The strict criterion was reached after processing a mean of 3837 syllables, with a range of 1380 to 8796 syllables. These values should be compared with 4536, which is the actual number of syllables read in the familiarization session of the Saffran et al. (1996b) experiment. By this time, all of our simulated participants had reached the first criterion, and 72% had reached the second criterion. These results make further comparisons with performance in the forced-choice test used by Saffran et al. (1996b), in which participants heard a pair of items including one word and one nonword on each trial, pointless. Indeed, simulated participants would outperform their actual counterparts, whatever the algorithm used to translate word knowledge into a performance level on the forced-choice test.

Beyond demonstrating the striking power of the principles underpinning PARSER, this first study revealed a point that is worthy of comment. The interest of PARSER, in our view, is that it accounts for a seemingly complex behavior by relying on general psychological principles rather than specific computational abilities. However, this interest could fade if, despite its commitment to these principles, PARSER required a memory capacity exceeding that which can be reasonably attributed to a human. To address the point, we considered the number of units that were simultaneously present in PS. Across all the simulations and the steps of learning, the maximum number of units able to shape perception in PS was 20. If one excludes the one-syllable units, this value falls to 9 units. If one considers, in addition, that several units are partially redundant (such as *pada* and *bupada*), such an amount appears fairly limited, quite compatible with the amount of information that human participants may have available in memory after a few minutes of practice with this kind of material.

However, PARSER failed to reproduce one aspect of the empirical data. Saffran et al. (1996b) examined whether performances differed as a function of the words. The words differed depending on the strengths of the transitional probabilities between their syllables. The authors reasoned that if participants' performance was based on the computation of transitional probabilities, the words with the higher transitional probabilities would be learned better than the words with the lower transitional probabilities. When splitting the six words into two sets according to this criterion, Saffran et al. indeed observed that the words with the higher mean transitional probabilities (*dutaba, pidabu,* and *tutibu,* from 1 to .75) led to better performance than the other words (*babupu, bupada,* and *patubi,* from .50 to .37). The mean scores were 79 and 72% correct responses, a difference that turned out to be significant. We examined whether our study reproduced the same result. To this end, the weight of the words after an arbitrary amount of training (9000 syllables) served as an index of learning. In accordance with Saffran et al.'s hypothesis, the word with the highest transitional probability (*dutaba,* with a mean transitional probability of 1) obtained, over the 100 simulations, a score significantly better than the mean score for the other five words (F (1,99) = 22.45, $p < .0001$). However, the scores for the other words were unrelated to their mean transitional probabilities, and the contrast between the two sets of three words as computed by Saffran et al. was not significant ($F < 1$).

The two words that were the more difficult to isolate for PARSER were *pidabu* and *tutibu.* Noteworthy, these words end in *bu,* which is the most frequent syllable in the language. This feature is a priori detrimental to the performance of the model, and we obtained direct confirmation of this contention. For instance, pairwise comparisons on simulated data revealed that *patubi* was learned significantly better than *tutibu.* However, this relation was reversed when the last syllable of these two words were inverted, the remainder of the material being unchanged: *tutibi* was then learned significantly better than *patubu.* The point is that the detrimental effect of the high frequency of *bu* on model's performance may have been masked in human participants, due to the introduction of another effect. One may speculate that *bu,* because of its repetition, becomes perceptually salient and serves as a natural anchor to parse the language into subunits. Suppose that a salient syllable tends to provoke the end of a percept: this would lead to participants performing better with *pidabu* and *tutibu,* in which *bu* is at the end of the words, than with *babupu* and *bupada,* which is the finding actually observed by Saffran et al. (1996b). The failure of PARSER to reproduce such an effect reveals a more general deficiency, which is PARSER's inability to take into account a number of statistical regularities that may be abtracted from a language. We will return to this point in the General Discussion.

Apart from this point, the results of Study 1 provided strong support for the validity of the principles underpinning PARSER. The model appears able to parse an unknown language into words after relatively little practice, while relying on quite elementary processes and limited memory capacities. In the two following studies, we explore whether PARSER is able to account for the other data reported by Saffran et al.

Study 2

In the Saffran et al. (1997) studies, the language was presented to participants while their primary task was to create an illustration with a coloring program. In the subsequent forced choice test, they selected the words over the nonwords significantly more often than would be expected by chance. On the one hand, the observation of learning in these incidental conditions appears consonant with our view of parsing as a by-product of the on-line processing of the items, without any superimposed analysis or computation. On the other, the task used by Saffran et al. (1997) was not only devised to prevent intentional analysis of the material; it was also intended to focus participant's attention away from the material. Our account, however, relies heavily on the properties of *attentional* processing. How can the possibility of learning in this context be encompassed within our account? Two remarks need to be made.

Firstly, the score observed under incidental conditions, although better than chance, was lower than the score obtained under full attention. Given that the material and the procedure were identical with that of Saffran et al. (1996b), except for the incidental learning condition, a direct comparison is possible. It appears that, with the same amount of training (one 21-min session), incidental conditions lead to 58.6% correct responding, whereas intentional conditions lead to 76%. A score of 58.6%

correct is surprisingly easy to achieve. In the Saffran et al. (1997) studies, participants were tested with 36 pairs of items, with each of the six words being paired with each of the six nonwords. Suppose that participants knew only one of the six words. This should lead to them correctly choosing the word over the nonword for 6 out of the 36 pairs, and to random responses for the 30 remaining pairs. This means that the final mean score would be 21 (6 + 30/2) correct responses out of 36, thus giving 58.33% correct responses. This value is hardly lower than the percentage above reported.

Secondly, the manipulation of attention devised by Saffran et al. (1997) was not stringent, to say the least. Participants were informed that the message displayed by the audiotape might affect their artistic creativity and that the experiment was aimed at investigating this influence. They were *not* instructed to ignore this source, and it is highly plausible that participants shifted attention toward the auditory message at least for some time within the 21-min sessions.

These observations led us to hypothesize that the reduced amount of attention paid to the language was sufficient to account for the moderate level of performance reported in Saffran et al. (1997). In order to obtain a more quantitative evaluation of this hypothesis, Study 2 was devised to estimate the proportion of items to which participants needed to attend to perform at the observed level.

To this end, only a fraction of the whole sequence of syllables was processed. In addition, a procedure was designed to simulate the attentional shifts of the participants. This procedure was needed because it is a priori easier for PARSER to process, say, a continuous run of 100 syllables than 10 times 10 syllables extracted randomly from a larger sample. The reason is that, in PARSER, correct coding of a word provides a starting point for the next percept that matches the beginning of the next word. This advantage is lost if attention is reoriented toward the language at random moments after a shift toward the coloring task. To take account of this point, the syllables were processed, in Study 2, in blocks of five percepts. After each block, a random number of syllables was skipped, so that the chance of beginning the next block at the beginning of a word was random, irrespective of the fact that the last percept of the prior block matched a word. The scores in the forced choice test were simulated through a simple model. For each test trial, the program selected the word over the nonword if the word was one individuated item in PS, provided this item had a weight greater than 1 (that is, if the unit was able to guide perception during the training phase). If the word was not in PS (or only as a component of a larger unit, or with a weight lower than 1), the choice was random.

In Experiment 1 of Saffran et al. (1997), participants were presented with 1800 words during the familiarization phase and were then tested with 36 pairs of items (each of the six words was paired with each of the six nonwords). The mean scores were 21.1 for adults and 21.3 for children, where chance is 18. PARSER provided a comparable score (21.34) while actually processing only 3% of the training material. The individual scores ($N = 100$) ranged from 13 to 31. These values are close to the actual data (as shown in Saffran et al., 1997, Fig. 1), although dispersion was slightly higher in simulated than in actual performance. In Saffran et al.'s Experiment 2, participants were presented with two sessions of 1800 words and were then tested as in Experiment 1. The mean scores were 26.3 for adults and 24.6 for children. PARSER provided mean scores of 25.9 and 27.02 with 4 and 5% of the material processed, respectively. Ranges of individual data tended again to slightly exceed the observed ranges. Thus, PARSER was able to simulate the actual mean scores while processing only a fragmentary part (about 3–5%) of the sequence presented to participants. This finding supports the idea that above chance performance in the incidental learning situation used by Saffran et al. is compatible with a framework grounded on the properties of attentional mechanisms. More generally, it suggests that continuous attention to the speech in standard training conditions may be unnecessary, a conclusion that makes our account even more plausible.

Study 3

The two prior studies used very same materials, a set of six trisyllabic words. In their experiments with 8-month-old infants, Saffran et al. (1996a) used only four trisyllabic words. For instance, the words used in one condition of their Experiment 2 were *pabiku, tibudo, golatu,* and *daropi.* Although the words changed as a function of groups and experiments, this is irrelevant for our purposes because all the syllables are equivalent for PARSER. However, in each case, and in contrast with the stimuli used in Studies 1 and 2, each syllable occurred only once in the four words presented to any given participant. Study 3 examined to what extent PARSER is able to parse this type of material.

Unexpectedly, PARSER, when run with the same parameters as in the prior studies, tended to perform worse with four words than with six words, at least when performance was evaluated by the criteria used above. It turned out that PS, even after extensive training, included a number of legal clauses (i.e., concatenations of a few words) with strong weights, which prevented the program from reaching even the loose criterion. The source of this difficulty was easy to detect: decreasing the number of words increased the probability of occurrence of any sequence of two or three words. As a consequence, legal clauses were progressively strengthened. In order to prevent this phenomenon, the following simulations were run after the forgetting rate was slightly increased, from 0.05 to 0.08. Note that this change is consonant with the common belief about the limited efficiency of infant memory with regard to adults (although it was not initially motivated by this consideration).

In a first step, we ran a set of simulations in order to examine when the criteria used above were reached. The loose criterion was defined by the fact that the four words are formed in PS and have the highest weights. This criterion was meet after processing a mean of 1077 syllables, with a range of 90 to 4908 syllables. As mentioned above, PS at this stage also includes other items, and notably part-words. The strict criterion was defined by the fact that the only residual items in PS are legal clauses, that is, strings composed of a few words. This criterion was reached after processing a mean of 1188 syllables, with a range of 156 to 4908 syllables. Although these values correspond to a fairly limited exposure to the material, they nevertheless exceed the actual practice received by infants. Indeed, infants were exposed to only 45 occurrences of each of the four trisyllabic words, that is, to 540 syllables.

In a second step, we examined the performance of PARSER when only 540 syllables were processed, in a way that may be easily related to infants' performance. In the Saffran et al. (1996a) experiments, the infants were presented with four items during the test. Two items were words in the language, whereas the two other items were nonwords. In Experiment 1, the nonwords were random three-syllable sequences of the syllables heard during the familiarization phase. In Experiment 2, the nonwords were three-syllable sequences that the infants had heard during the familiarization phase, but that crossed the word boundaries. For instance, for the participants who perceived *pabiku, tibudo, golatu,* and *daropi* during familiarization, the test items were *pabiku, tibudo, tudaro,* and *pigola.* Each item was repeated, and the infants controlled their listening duration. Nonwords elicited longer listening times than words. Of course, we have no available model of performance for simulating the listening duration of infants; however, examining whether the word items are more often in PS than the nonword items provides a clue about the ability of PARSER to account for infants' behavior.

Over the 100 simulations, 52 had the two words involved in the test in PS (with a weight greater than 1), and 93 had at least one of the two words after processing 540 syllables. These values fell to 23 and 65, respectively, if one considered only the four items in PS that had the strongest weights. Considering only the most highly weighted item in PS, there were still 23 simulations for which this item matched one of the two words displayed during the test. In contrast, the nonword items spanning word boundaries that were used in the test phase of Experiment 2 (the nonword items used in Experiment 1 could not be in PS, because they were composed of syllables which were never seen in immediate succession) were never found in PS, even when no limiting conditions on the weights were imposed. These findings allow only indirect conclusions, due to the arbitrariness of translating a weight in PS into a listening duration. However, they show that after processing the same number of syllables as the infants, PARSER had extracted a high proportion of the words and none of the nonwords used during the test, even when the sequences forming the nonwords were displayed during familiarization. The contrast between words and nonwords persisted even if the analysis was restricted to the most highly weighted item in PS.

Study 4

All the experiments published to date by Saffran and co-workers used trisyllabic words. This feature induced a regularity that may have helped participants to parse the sequence into words. PARSER is obviously unable to draw information from this regularity. However, it may be argued that its parameters, and notably the fact that each attentional span is fixed to between one and three primitives, makes the task of discovering three-syllable words especially easy. By the same token, the association-based, incremental mode of word building may suggest that short words would remain undetected. In order to address this issue, we ran a last set of simulations with words of varied length. The words displayed during the training phase were *bu*, *bapa*, *dutabu*, *patubi*, *pidabupu*, and *tutibudaba*. These words included the same syllables as those used by Saffran et al. (1996b, 1997), but in such a way that the length of words went from one to five syllables. It should also be noted that the one-syllable word (*bu*) was the most frequent syllable of the language, a feature that a priori increased the difficulty of the parsing task.

The simulations were run with the same parameters as for Studies 1 and 2. The loose criterion was reached after processing a mean of 1910 syllables, with a range of 613 to 4782 syllables. The strict criterion was reached after processing a mean of 3338 syllables, with a range of 1184 to 8538 syllables. Fine comparisons with the results obtained with three-syllable words make little sense because the precise values may depend on the composition of the material and on the choice of program parameters. However, given that the values obtained here are lower than the values collected in Study 1 with three-syllable words, a conservative conclusion is that using words of different lengths has no detrimental effect on the performance of PARSER.

GENERAL DISCUSSION

The studies conducted by Saffran and co-workers (1996a, b, 1997) revealed that people were able to parse an artificial language into words even though the language was displayed without any physical word boundaries or other prosodic cues. This performance, according to the authors, implies that people draw an inference about the location of word boundaries, based on the fact that transitional probabilities are higher for word-internal than for word-external pairs of syllables. We propose here a strikingly different interpretation. Our account assumes that the material is mandatorily perceived as a succession of small and disjunctive chunks composed of a few primitives. This characteristic is thought to be inherent in the attentional processing of ongoing information. When a chunk is repeatedly perceived, its components are associated and form a new representational unit as an automatic by-product of the joint attentional processing of the components. The units of the language initially emerge thanks to a sort of natural selection process: among all the units that are created, only those matching the words (or parts of words) are sufficiently repeated to resist forgetting and interference. These initial representational units in turn become able to guide perception and to enter as components of other percepts, and this process continues recursively. This ensures that the system converges rapidly toward the words.

These principles were implemented in PARSER in order to assess their explanatory power. A comparison of our results with those of Saffran et al. showed that PARSER, as a rule, learns at least as well as human participants. In Study 1, PARSER extracted the words without any errors before the end of the training session experienced by the participants in Saffran et al. (1996b). Study 2 showed that PARSER was able to simulate the performance of people trained under incidental conditions (Saffran et al., 1997) after processing only 3 to 5% of the material. Study 3 demonstrated that infants' performance reported in Saffran et al. (1996a) could be reproduced even if one considers only a very few items among the most highly weighted items learned by PARSER. Thus, PARSER provided a good overall match with the performance observed in the Saffran et al. experiments. Where the simulations exceed human performance, the addition of some noise at different steps of the model would appear to be an obvious remedy.[3] Finally, Study 4 showed that the good performance of PARSER was not limited to the three-syllable words used in all of the Saffran et al. studies, but also extended to a language composed of one- to five-syllable words.

Thus, PARSER proved to be able to segment linguistic materials identical or similar to those used in the Saffran et al. experiments, without relying on sophisticated computational devices or extensive memory capacities. However, the model's ability to reproduce the experimental results of Saffran et al. does not establish its relevance with regard to acquiring natural language in real-life conditions. The material used in the experiments of Saffran and co-workers, as well as the conditions in which this material was presented to the participants, ignored a number of important aspects of natural language. Some differences were inherent in any laboratory analog of a lifesize problem. Natural language acquisition does not consist in identifying six words used again and again in a few minutes, but many thousands of words distributed over years. Are the principles underlying PARSER general enough to be easily applied to such different complexity and timing scales? Other differences between the material used by Saffran and co-workers and the natural language were introduced to meet the requirement of authors' research strategy. In natural acquisition, children are exposed to short utterances separated by clear pauses, including variations in pitch and stress. In addition, any natural language contains a number of phonotactic regularities. Recent psycholinguistic research has given evidence that these features may be used as cues for word boundaries during language acquisition (see Brent & Cartwright, 1996; Christiansen, Allen, & Seidenberg, in press; Jusczyk, 1997; McDonald, 1997). All of these features were removed from the situation of Saffran et al. Hence, the present version of PARSER says nothing about these aspects. We must address the way in which the multiple cues are integrated to guide speech segmentation if we wish to propose PARSER as a kernel of a realistic model of language acquisition. These different issues will be discussed in turn.

Applying PARSER to Lifesize Problems

As we have mentioned, PARSER works thanks to the interaction between one property of the language and a few properties of the human processing system. Are there any reasons to believe that this interaction occurs only with the simplistic language used by Saffran and co-workers? The target property of the language, namely that the probability of repeatedly selecting the same group of syllables by chance is higher if these syllables form within-word rather than between-words components, is obviously shared by the artificial Saffran et al. material and by any natural language. Likewise, the properties of the processing system on which PARSER relies are very general. For instance, one fundamental assumption of the model is that a cognitive unit is forgotten when not repeated and strengthened with repetition. This assumption may be taken for granted irrespective of whether the process occurs in the few minutes of an experimental session or across larger time scales, in keeping with a

[3] As suggested by an anonymous reviewer, it is also possible that the overachievement of the model is not due to PARSER itself, but rather to the way in which performance in the final tests of Saffran et al. (1996a, 1996b, 1997) was estimated from the word knowledge provided by PARSER. We assumed no memory decrement during this phase. The model may outperform humans because the participants in final forced-choice test of Saffran et al. may have experienced interference from the repeated presentation of nonwords and part-words.

long-standing tradition of research on the laws of memory and associative learning. In consequence, PARSER's principles seem to be relevant to natural as well as to artificial language. Briefly stated, the generality of PARSER is ensured by both the generality of the behavioral laws (e.g., only repeated units shape long-lasting representations) and the generality of the language property (the most repeated units are the words) on which it relies.

However, despite the theoretical relevance of its principles, the application of the present version of PARSER to the acquisition of natural language is far from trivial. Adults modify their language to some extent when interacting with children. Child-directed language (‘‘motherese’’) is simpler than adult language, and speech addressed to preverbal infants is simpler still than speech addressed to older children (e.g., Aslin, Woodward, LaMendola, & Bever, 1996). Therefore, we have to consider two issues in turn, pertaining, respectively, to the ability of PARSER to deal with motherese and to the value of PARSER as a model of learning with an input of increasing complexity.

To address the first issue, let us consider the Korman (1984) corpus, which consists of speech directed to infants before 16 weeks of age. This corpus contains 1888 types of different words. If one leaves out onomatopoetic words and interjections that often occur in isolation, the corpus still contains more than 1000 words (Christiansen et al., in press). This value is lower than the tens of thousands of words composing adults' language, but is arguably still more unlike the four to six words used by Saffran and co-workers. The problem does not stem from the number of words per se, because there is no theoretical reason to limit the size of the percept shaper, the memory store, or mental lexicon of PARSER. However, it could be argued that PARSER would work poorly while trying to segment motherese, because the rate of forgetting used in the present simulations is too high to allow a given unit to be perceived again before it is dropped from the percept shaper.

One possible counterargument is that motherese is quite redundant. The Korman (1984) corpus has a type-token ratio of .05, indicating a considerable amount of repetition. For comparison, the type-token ratio in the Saffran et al. (1996b) experiments with 8-month-old infants was .022. The rate of repetition is higher in the Saffran et al. material, but the difference is not as striking as could be anticipated. Unfortunately, the comparison is only partially relevant, because, even if one assumes a fixed *rate* of repetition, the lag between two occurrences of the same word would still depend on the size of the corpus, with a larger corpus having a longer mean lag. Thus, we suspect that the present incarnation of PARSER would fail to parse the Korman corpus because the chance for a unit to be perceived again before disappearing from the percept shaper would be very low.

An obvious remedy might be to reduce the forgetting rate. However, this would probably generate an unrealistic number of strongly weighted units in the percept shaper for a simple artificial language, and adjusting a parameter as a function of the to-be-learned material appears unsuitable. Fortunately, there is a more elegant solution. In the present version of PARSER, forgetting was simulated through a linear decrement. This may be a reasonable approximation when only short time intervals are considered, as in the present experimental settings. However, there is evidence that the forgetting curve fits only moderately well with a linear trend (e.g., Rubin & Wenzel, 1996), especially when the study–test interval increases. A more realistic power function, for instance, would make it possible to combine a rapid initial decay, which may be appropriate for most experimental situations, with the long term persistence of residual traces in memory, which is needed for the acquisition of a language under natural conditions. Such a modification of the forgetting function would not be a poorly motivated ad-hoc adjustment of parameters, but instead would embed PARSER even more solidly within our knowledge about human memory.

Although the relevance of PARSER for dealing with motherese is of primary concern for language acquisition, it must be kept in mind that the final outcome of the process is adults' natural language mastery. This raises the second issue, regarding the value of PARSER as a model of learning with an incremental input. PARSER appears especially well-suited for such a task. Indeed, in PARSER, any new perceived unit does not affect the nature of the units already present in the percept shaper. It does affect the weight of similar units through the process of retroactive interference, but in a quite limited way. The effect of interference becomes practically negligible for a given unit once this unit has been strengthened a few times. This ensures stability of the system in the face of new input. Moreover, in PARSER, the units the processing system has already found are used to segment new utterances, and hence to discover new units. This property is a consequence of one of the fundamental psychological principles in which PARSER is rooted, namely that internal representations guide perception. Thus, when the system is faced with incremental input, its constituent units are not only left almost intact, but help with the relevant coding of this input.

These remarks about PARSER's power are mostly speculative. These speculations need to be empirically supported by running simulations on actual corpuses. As mentioned above, this entails some modifications to make the program more realistic, at least with regard to the forgetting function. We now examine whether other modifications would be possible to take into account some of the multiple cues for word boundaries available in natural language.

The Integration of Other Cues for Segmenting Speech

Earlier, we mentioned some of the prosodic and phonological cues that children may use to segment the speech stream into words (for recent review, see Brent & Cartwright, 1996; McDonald, 1997; Jusczyk, 1997). All of these cues were intentionally removed from the material used by Saffran et al. Participants' achievement in these studies, as well as the performance of PARSER, suggest that such cues are not strictly necessary to correctly segment linguistic utterances. This does not mean that the prosodic and phonological cues do not intervene in the segmentation process. It is beyond the scope of this paper to make PARSER able to integrate some of these aspects, as for instance, Christiansen et al. (in press) were recently able to do for Elman's (1990) simple recurrent network. However, it is worth considering whether such an integration would be possible, and how an improved version of the model might work.

Let us first consider the pauses in the speech flow. Exploitation of the pauses could be implemented in the program step in which the size of the next percept is selected (Fig. 1, step a). In the current version of the model, the number of units processed in one attentional focus is randomly determined. We postulated that this procedure was appropriate for reproducing the effects of the various and uncontrolled factors which modify the boundaries of the momentary attentional window, such as the listener's state of vigilance, when a participant is exposed to a new language in Saffran and collaborators' experiments. However, in real-world settings, pauses provide natural cues for segmenting the speech flow from its very beginning. Although the information is insufficient for full segmentation, it may be quite useful for children given that child-directed language is characterized by very short utterances separated by clear pauses. Incorporating the information provided by the pauses into PARSER is straightforward: we simply need to constrain selection of the number of units perceived in one attentional focus such that the content of an attentional focus does not straddle pauses. This would make the model more powerful, although the exact amount of improvement is a matter for further simulation studies.

The exploitation of additional prosodic and phonological features correlated with word boundaries (e.g., rhythm, allophonic variation) raises a more complicated problem. When the function of these features is known, their use could be implemented in PARSER in the same way as pauses, that is, as additional constraints on the boundaries of the successive percepts. However, unlike pauses, other prosodic and phonological cues do not provide a natural indication for segmentation. Their function in a particular language needs to be inferred by the listener. PARSER had no mechanisms to make such inferences. Because use of this information may be increasingly important with a natural language as expertise progresses, this inability reveals an important limitation on the generality of PARSER. However, this does not hamper the utility of PARSER even with regard to the exploitation of these cues. We pointed out in the introduction that basing word segmentation exclusively on prosodic and phonological cues would be circular, given that discovering the statistical regularities associated with the word-level organization of the language implies prior knowledge of at least some words. The mechanisms involved in PARSER could allow the formation of an initial knowledge base from which prosodic and phonological regularities could be subsequently inferred and exploited.

Conclusion

We have presented a new account of the ability to extract words from a continuous flow of speech that lacks prosodic cues, as evidenced by Saffran and co-workers in adults, children, and infants. Our account relies on general principles of memory and associative learning and requires no computational abilities besides those involved in any memory-based behavior. The achievement of PARSER, a computer model implementing these principles, supported our account of word segmentation, at least when applied to an artificial languages identical or similar to those used by Saffran and co-workers. Admittedly, some aspects of the present version of PARSER, notably the forgetting function, appear ill-suited for dealing with natural language acquisition. These aspects ought to be modified in further versions of the model, in order to assess PARSER’s ability to parse into words a natural language corpus. Our claim is that generalization to the acquisition of natural language can be envisioned with reasonable optimism, because the psychological principles, as well as the property of language that triggers the efficiency of these principles, extend quite well from laboratory conditions to real-life settings.

REFERENCES

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan and K. Demuth (Eds.), *Signal to Syntax* (pp. 117–134). Mahvah, NJ: Erlbaum.

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

Case, R. (1993). Theories of learning and theories of development. *Educational Psychologist,* **28,** 219–233.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (in press). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes.*

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

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language,* **36,** 202–225.

Eimas, P. D. (1997). Infant speech perception: processing characteristics, representational units, and the learning of words. In R. L. Goldstone, P. Schyns, and D. E. Medin (Eds.), *The psychology of learning and motivation,* (Vol. 36, pp. 127–169). San Diego: Academic Press.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science,* **14,** 179–211.

Finney, S., Protopapas, A. & Eimas, P. D. (1996). Attentional allocation to syllables in American English. *Journal of Memory and Language,* **35,** 893–909.

Fischer, K., & Granott, N. (1995). Beyond one-dimensional change: Parallel, concurrent, socially distributed processes in learning and development. *Human Development,* **38,** 302–314.

Goldstone, R. L., Schyns, P., & Medin, D. E. (1997). Learning to bridge between perception and cognition. In R. L. Goldstone, P. Schyns, & D. E. Medin (Eds.), *The psychology of learning and motivation,* (Vol. 36, pp. 1–14). San Diego: Academic Press.

Jusczyk, P. W. (1997). *The discovery of spoken language.* Cambridge, MA: MIT Press.

Jusczyk, P. W., Kennedy, L. J., & Jusczyk, A. M. (1995). Young infants’ retention of information about syllables. *Infant Behavior and Development,* **18,** 27–42.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science.* Cambridge, MA: Bradford/MIT Press.

Korman, M. (1984). Adaptive aspects of maternal vocalizations in differing contexts at ten weeks. *First Language,* **5,** 44–45.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review,* **82,** 276–298.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology,* **10,** 29–63.

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

McDonald, J. L. (1997). Language acquisition: The acquisition of linguistic structure in normal and special populations. *Annual Review of Psychology,* **48,** 215–241.

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

Perruchet, P., & Gallego, J. (1997). A subjective unit formation account of implicit learning. In D. Berry (Ed.), *How implicit is implicit learning* (pp. 124–161). Oxford: Oxford University Press.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General,* **119,** 264–275.

Perruchet, P., & Vinter, A. (1998). Learning and development: The implicit knowledge assumption reconsidered. In M. Stadler and P. Frensch (Eds.), *Handbook of implicit learning* (pp. 495–531). Thousand Oaks, CA: Sage.

Perruchet, P., Vinter, A., & Gallego, J. (1997). Implicit learning shapes new conscious percepts and representations. *Psychonomic Bulletin and Review,* **4,** 43–48.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development.* Chicago: University of Chicago Press.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review,* **103,** 734–760.

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

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language,* **35,** 606–621.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning. *Psychological Science,* **8,** 101–105.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts, *Behavioral and Brain Sciences,* **21,** 1–53.

Servan-Schreiber, D., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **16,** 592–608.

Vinter, A., & Perruchet, P. (1997). Relational problems are not fully solved by a temporal sequence of statistical learning episodes. *Behavioral and Brain Sciences,* **20,** 82.

(Received November 17, 1997)
(Revision received March 30, 1998)