Learning Nonadjacent Dependencies: No Need for Algebraic-Like Computations

Pierre Perruchet, Michael D. Tyler, Nadine Galland, and Ronald Peereman
University of Bourgogne

Is it possible to learn the relation between 2 nonadjacent events? M. Peña, L. L. Bonatti, M. Nespor, and J. Mehler (2002) claimed this to be possible, but only in conditions suggesting the involvement of algebraic-like computations. The present article reports simulation studies and experimental data showing that the observations on which Peña et al. grounded their reasoning were flawed by deep methodological inadequacies. When the invalid data are set aside, the available evidence fits exactly with the predictions of a theory relying on ubiquitous associative mechanisms. Because nonadjacent dependencies are frequent in natural language, this reappraisal has far-reaching implications for the current debate on the need for rule-based computations in human adaptation to complex structures.

The idea that most higher cognitive activities, especially language comprehension and production, are based on abstract, rule-based operations on symbolic contents is one of the cornerstones of the mainstream tradition in cognitive psychology. However, over the last 2 decades or so, an alternative view has gained increasing influence. This alternative conception is rooted in the traditional associative view of mind, but its recent upsurge is essentially related to the growth of connectionist modeling (e.g., McClelland & Rumelhart, 1986). Indeed, connectionist studies have shown that certain activities that were once thought of as straightforward evidence for rule-based computations can be simulated by models that rely only on associative mechanisms. This issue has crystallized around two broad conceptions of the mind, often thought of today as an opposition between those who advocate the need for assuming algebraic-like computations (e.g., Marcus, Vijayan, Rao, & Vishton, 1999; Pinker, 1997) and the proponents of statistical/distributional approaches (e.g., Redington & Chater, 1998; Seidenberg & MacDonald, 1999). The English past tense has been a focus for this debate from its outset, and the number of papers pertaining to this issue over the last few years (e.g., Pinker & Ullman, 2002; Ramscar, 2002) suggests that it is still unsettled despite a considerable amount of research effort.

A newcomer in this lively debate is the learning of nonadjacent (or remote) dependencies. The major part of the traditional literature on associative learning has dealt with relations between adjacent events. This is true both for the domain of animal conditioning and for studies on paired-associate learning in humans. In both cases, the items to be associated are displayed in close temporal or spatial proximity. The same is true for the more recent studies on implicit learning (e.g. Stadler & Frensch, 1998). Looking at a standard flowchart of a finite-state grammar commonly used in artificial grammar studies is sufficient to show that relations are built between contiguous elements.[1] Several studies have shown that those adjacent relations were far more relevant for linguistic structure than researchers had claimed in the past. For instance, Redington, Chater, and Finch (1998) showed that highly local context provided a considerable amount of information about the syntactic categories of words. However, it appears that linguistic structures also embed remote dependencies. That is to say, a relation exists between *A* and *C* irrespective of the intervening events (hereafter, this structure is referred to as *AXC*, where *X* stands for a variable event statistically independent from both *A* and *C*). Such relations are found at different levels, from the subsyllabic level (e.g., the short vs. long pronunciations of vowels according to the presence of a silent *e* ending, irrespective of the intermediary consonant, as in *cap* vs. *cape* and *car* vs. *care*; Stanback, 1992) to morphosyntactic relationships (e.g., between auxiliaries and inflectional morphemes, as in "*is* writi*ng*," irrespective of the verb stem) and hierarchical structures (e.g., in center-embedded sentences, e.g., "The rat the cat ate stole the cheese"). If associative mechanisms turn out to be able to learn from nonadjacent dependencies, this will undoubtedly increase their potential participation in language acquisition.

Pierre Perruchet, Michael D. Tyler, Nadine Galland, and Ronald Peereman, Laboratory for the Study of Learning and Development/National Center for Scientific Research (LEAD/CNRS), University of Bourgogne, Dijon, France.

Correspondence concerning this article should be addressed to Pierre Perruchet, Université de Bourgogne, LEAD/CNRS, Pôle AAFE, Esplanade Erasme, 21000, Dijon, France. E-mail: pierre.perruchet@u-bourgogne.fr

---

[1] This does not mean that relations between remote events have been totally ignored. For instance, the possibility of detecting second-order dependencies in implicit learning settings has been the topic of a number of investigations (e.g., Reed & Johnson, 1994). However, in each case, the relations between distant events are not considered independently from the intervening events. For instance, given the sequence *ABC*, the *A–C* relation is assessed indirectly, through improvement in the prediction of *C* allowed by the consideration of *AB* as a whole.

The Peña, Bonatti, Nespor, and Mehler (2002) Study

In light of these potential implications, a few studies have recently investigated the acquisition of remote dependencies (Cleeremans, 1993, Chapter 3; Gomez, 2002; Newport & Aslin, 2004; Santelmann & Jusczyk, 1998). The present article focuses on a study published in a recent issue of *Science* (Peña et al., 2002). The authors presented a set of new experiments that, according to them, provide a case against the conclusion that a statistical or distributional approach is able to account for the discovery of nonadjacent dependencies. This form of learning would require some kind of algebraic-like computations.

In their experiments, Peña et al. (2002) presented a stream of syllables, built from the concatenation of trisyllabic artificial words that were defined by the relation between the first and the last syllable of each word. There were three families of words ($A1XC1$, $A2XC2$, and $A3XC3$), and within each family, the first syllable (A) predicted the last syllable (C) with a probability of 1. The intermediary syllables, Xs, could be one of three other syllables, with the same pool of Xs serving for all the families. In a subsequent test phase, participants were presented with pairs of trisyllabic items, and they were asked to judge which item from each pair sounded more like a word of the imaginary language. In Experiment 1, one item of each pair was a word (e.g., $A2XC2$), and the other item, which they called a *part word*, was formed by the end of one word and the beginning of another (e.g., $XC1A2$, or $C2A3X$). In those conditions, participants selected significantly more words than part words, a result that, according to the authors, provides evidence that participants had detected the remote contingencies. In Experiment 2, to ensure that *generalization* occurred to all members of each family, the authors did not take the X syllables entering into the test items from the pool used during familiarization. Instead, they took these syllables from the list of A or C syllables (e.g., $A1A3C1$, or $A3C2C3$). Peña et al. called these test items the *rule words*. The rule words were not selected more often than the part words, which suggests that learning now failed to occur.

According to Peña et al. (2002), this pattern of results attests that statistical mechanisms are sufficient to allow learning of remote dependencies when the test contrasts part words and words but are unable to ensure generalization to rule words. To explore the hypothesis that rulelike mechanisms are required in this latter case, Peña et al. introduced what they called a subliminal gap of 25-ms duration between the words during the familiarization phase (Experiments 3 and 5). "The role of silent gaps," Peña et al. asserted, "is to make the stream more similar to natural language" (p. 606). So doing, they trigger computational processes oriented toward the discovery of the speech structure, which are entirely different from those involved in the discovery of statistical pattern. Although this condition was "subjectively very similar" (p. 606) to the no-pause situation according to Peña et al., learning now occurred. In Experiment 3, with a 10-min-long familiarization phase, participants chose the rule words over the part words for 69.6% of the pairs ($p < .0005$). The authors concluded that, when the 25-ms pauses were present, participants "spontaneously formulated an implicit grammatical-like generalization that corresponds to the structure of the represented items" (p. 606). In Experiment 5, the initial exposure was reduced to 2 min, and participants still chose rule words over part words for 67.1% of the pairs ($p < .0005$). The fact that sensitivity to nonadjacent dependencies appeared after a very limited exposure to the speech flow was taken as a supplementary indication according to which the recruited processes are not incremental/statistical in nature but rely on grammatical-like computations.

A Logical Drawback

The line of reasoning used by Peña et al. (2002) could be questioned on several points. For instance, why subliminal pauses should trigger algebraic-based mechanisms normally devoted to natural speech is not clear, in view of the fact that there are no pauses between words in normal speech (e.g., Cole & Jakimik, 1980). Also, the proposal according to which rapid learning attests to rule inference is debatable, as we later show. However, let us provisionally take these assumptions for granted. A major logical pitfall remains.

The problem originates from the test designed for measuring the learning of remote contingencies. As stated above, the authors compared words (in Experiment 1) or rule words (in the other experiments) with part words—that is to say, with trisyllabic items spanning word boundaries—following from studies investigating word segmentation (e.g., Saffran, Newport, & Aslin, 1996). The authors' explicit reasoning was that if participants discovered the A–C relationships, then they would organize the stream into AXC words. However, a stronger condition is in fact required. Inferring the discovery of the A–C relationships from the segmentation into AXC words requires that remote dependencies are not only one source but the exclusive source of information exploited by participants to parse the speech flow into the AXC words. If participants also rely on other cues to segment the auditory string, then measuring the learning of remote dependencies through its effect on word segmentation becomes clearly unwarranted. Now, it appears that segmentation may be grounded in a variety of information sources, whether they are statistical, phonotactic, or prosodic. Of special relevance in this context is the fact that Peña et al. (2002) introduced a gap of 25-ms duration between the words during the familiarization phase in their Experiments 3 and 5. They acknowledged that "the gaps contained in the stream may help participants to segment" because they provide "explicit bracketing cues" (p. 606). However, the authors failed to note that in introducing this gap, they violated the postulate underpinning their test, namely that remote dependency learning is the exclusive cue for word segmentation.

The Present Study

In the first part of our study, we explore what cues other than the distant dependencies were available to learners in Peña et al.'s (2002) experiments. For the sake of clarity, we distinguish between two nonexclusive categories of cues. Some of them may be linked to the relational structure of the syllables in Peña et al.'s language. Indeed, the presence of distant relationships, coupled with the order constraints regulating the succession of syllables, may have introduced some unwanted statistical regularities. Because these regularities need to be learned during the familiarization phase to be influential on segmentation, we call this first kind of cues *training dependent*. By contrast, other cues are directly exploitable by the participants, because they are related to natural perceptual biases. In the context of artificial grammar learning studies, it has been shown that participants are sensitive to a large

number of features before any training with the artificial language (Reber & Perruchet, 2003). These biases may be acquired through the participants' experience with their natural language, but, because they are efficient from the outset in any experimental settings, we call them *training independent*. We investigate the role of training dependent and training independent factors in two simulation studies and one experiment specifically designed to tease apart each category of factors. We show that both categories of factors are able to account for word segmentation.

The second part of our study is aimed at evaluating whether and to what degree the above-mentioned factors are effective. Our target is Peña et al.'s (2002) Experiment 5, in which participants were shown to learn remote dependencies after a very short exposure to the speech stream. In Experiment 2, we used all of the study materials and some of the test items from Peña et al., but additional test items were designed to provide a methodologically sound measure of remote dependency learning. This measure was borrowed from earlier studies (e.g., Gomez, 2002). For the sake of simplicity, let us consider a case in which participants are presented with only two families of words: $A1XC1$ and $A2XC2$. Because the question of interest is whether participants have learned the dependencies between $A1$ and $C1$, on the one hand, and between $A2$ and $C2$, on the other, relevant control items are $A1XC2$ and $A2XC1$ (hereafter, we call these items *scrambled words*). If participants consistently chose rule words over scrambled words during the test, then it could be inferred that they had learned the genuine dependency relations. Furthermore, because participants have to choose on each trial between two items sharing the AXC pattern, this measure is not confounded with the participants' ability to segment the speech flow into the AXC units and therefore avoids the problem of assessing whether correct segmentation and remote dependency learning are causally related.

Finally, we discuss the implications of our results, showing that, when invalid data are set aside, the available evidence fits well with the predictions of a theory relying on simple and ubiquitous associative mechanisms.

Simulation Studies

Our first objective was to examine whether the materials displayed to participants embedded statistical cues for segmentation other than the remote dependencies between A and C. If that were the case, then using participants' word recognition ability as a measure of remote dependency learning would be misguiding. To meet this objective, we entered Peña et al.'s (2002) language into Parser (Perruchet & Vinter, 1998), a model based on elementary principles of associative learning that, by construction, is (a) unable to exploit nonadjacent statistical dependencies and (b) immune to any a priori knowledge and biases and hence insensitive to training-independent factors. If Parser achieves segmentation, then the input must contain statistical information for segmentation apart from nonadjacent dependencies.

Parser was devised to account for the results of Saffran, Newport, and Aslin (1996), in which the relevant relations are between adjacent syllables. Parser mimics the phenomenal experience of the learner. When people are confronted with material consisting of a succession of elements, each of them matching some of their processing primitives, it is assumed that they segment this material into small and disjunctive parts consisting of a small number of adjacent primitives. In Parser, each part is composed of one to

three processing primitives (the number is determined randomly for each percept). The primitives forming a chunk—that is, those that are perceived within one attentional focus as a consequence of their experienced temporal proximity—tend to pool together and form a new primitive for the system. As a consequence, they can enter as a unitary component into a new chunk in a further processing step. This explains why the phenomenal experience changes with practice. But why do the initial primitives evolve into a small number of words instead of innumerable irrelevant processing units?

The reason lies in the properties of the human processing system. The future of the chunk that forms a conscious episode depends on ubiquitous laws of associative learning and memory. If the same experience does not reoccur within some temporal lag, the possibility that a chunk will act as a processing primitive rapidly vanishes, as a consequence of both natural decay and interference from the processing of similar material. The chunks evolve into primitives only if they are repeated. Thus, some primitives emerge through a natural selection process, because forgetting and interference lead the human processing system to select the repeated parts from all of those generated by the initial, presumably mostly irrelevant, chunking of the material. The relevance of this phenomenon becomes clear when viewed in relation to a property inherent to any language. If the speech signal is segmented into small parts on a random basis, these parts have a greater chance of being repeated if they form a word, or part of a word, than if they straddle word boundaries. Consequently, the primitives that emerge from the natural selection because of forgetting and interference are more likely to match a word, or a part of a word, than a between-words segment. Technical details about Parser are provided in Perruchet and Vinter (1998), and an online presentation of the model is available on the Internet (http://www .u-bourgogne.fr/LEAD/people/perruchet/SOC.html).

The most crucial point is that all the representations created by Parser are composed of adjacent elements. Thus, Parser is, by construction, unable to exploit remote dependencies, as they were defined above. To explore whether Peña et al. (2002) actually demonstrated that participants exploit remote dependencies, we entered into Parser the material heard by participants in Peña et al.'s experiments. If Parser is successful, then the Peña et al. procedure is flawed.

Study 1

The material was generated exactly as in Experiments 1 to 5 of Peña et al. (2002). For instance, all of the word order constraints were respected. The number of words entered during the study phase was identical to the number of words shown in Peña et al.'s Experiment 1, which was our main target. The version of Parser was identical to that used in Perruchet and Vinter (1998), Perruchet, Vinter, Pacteau, and Gallego (2002), and Perruchet and Peereman (2004), except that the parameters indexing forgetting and interference were set to .005 and .001, respectively (these parameters need to be adjusted to the material and to the number of different words composing the stream). To simulate the results of the test used in Peña et al.'s Experiment 1, we compared the weight of the word and the weight of the part word of each pair and chose the item with the highest weight (if a word or a part word was not in the model's lexicon, we considered its weight to be zero). Because Peña et al.'s experiments involved 14 participants,

we ran 14 independent simulations to make the power of the statistical test identical to theirs. The mean score was 70.28% ($SE = 6.22$), a value that is significantly above chance, $t(13) = 3.26$, $p = .0062$, and close to the scores obtained by the actual participants in Peña et al.'s Experiment 1 (73.3%). Thus, Parser discovered the words without exploiting remote dependencies.

This performance is all but mysterious. Indeed, a close examination of the materials of Peña et al.'s (2002) experiments shows that words are twice as frequent as part words in the speech flow displayed during the familiarization phase. Parser, presumably, was sensitive to this feature, and, of course, actual participants of Peña et al.'s experiments may have proceeded in the same way.

Study 2

Peña et al. (2002) were at least partially aware of the potential confound highlighted by the results of Study 1. They observed that "part-words occur less frequently than words" and reasoned that "the outcome of Experiment 1 might be due to absolute frequency of syllable co-occurrence, rather than to distant transitional probabilities" (p. 607, Endnote 16). To disentangle the two accounts, they designed new materials in which one word from each family occurred two times more frequently than the other two words of the same family and then used only the less frequent items during the test. With this new method, the words and the part words composing each pair used during the test had exactly the same frequency of occurrence during the familiarization phase. Peña et al. reported that participants still chose the words for 70.5% of the pairs. They concluded that "it is not the absolute frequency of occurrence of the trisyllabic items that directs participants' choices, but the distant transitional probability relations among their syllables" (p. 607).

Again, we used the material played to participants as input for Parser, using exactly the same method and parameters as for Study 1, and Parser also learned the words. Words were chosen against part words for 68.75% ($SE = 5.43$) of the test pairs, $t(13) = 3.31$, $p = .0057$, even though words and part words had the same frequency of occurrence during the familiarization phase. A first possible explanation is that making frequency equal allows for other statistical cues for segmentation in the input. For instance, it is possible that transitional probabilities or contingencies differ according to whether intra- or interword syllables are considered, and Parser has been shown to be sensitive to these statistical measures of association (Perruchet & Peereman, 2004). However, this does not seem to be the case: The measures of association we examined could not be used to discover the words. The correct explanation lies elsewhere. In a first stage of learning, Parser naturally discovers the most frequent words—that is to say, two out of the three words in each family—by exploiting the distributional properties of the material, as in Saffran, Newport, and Aslin's (1996) language. The question is, Why does Parser find out the remaining third of the words, instead of the part words, which have the same distributional properties? Parser exploits here a property that is also exploited in other nonconnectionist models of word segmentation, such as the Brent and Cartwright (1996) model. Given that (a) the units are discrete and disjunctive and (b) perception is driven by the internal representations, the first words to be discovered guide the residual segmentation of the speech. Thus, unit extraction depends on the overall consistency of the segmentation. Parser discovered the most frequent words first, and

this knowledge base provided strong constraints on the segmentation of the adjacent material. The phenomenon at hand may be captured in everyday experience: An unknown word is immediately perceived as a unit when the surrounding words are familiar units.

As an aside, it is worth noting that this property of Parser expands considerably its relevance for natural language, because intraword constituents in natural language are more closely associated than interword constituents, as in Saffran and collaborators' artificial language (e.g., Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996), but only on average (e.g., the syllables /er/ and /kon/, present in the expression "air conditioned," are certainly more closely associated than those composing a number of rare words). We discuss more thoroughly this characteristic of Parser elsewhere (Perruchet & Peereman, 2004).

For the present concern, our conclusion is that the materials played to participants in Peña et al.'s (2002) experiments included statistical cues to parse the speech flow into words, independently of the A–C relationship, which the authors assumed to be the only source of relevant information. Note that our simulations were limited to two experiments, because only those experiments provided information about word segmentation comparable to the information Parser provides. In those cases, the performance of Parser, which is by construction insensitive to remote dependencies, was close to that of actual participants. However, the relevance of these results is by no means limited to these two experiments, because Peña et al. used the same sequences of words during the study phase of all their experiments (with the exceptions of a few experiments reported in footnotes, such as the experiment reanalyzed in our Study 2). Thus, it appears that, irrespective of whether the final test was designed to reveal this knowledge, the statistical information available in the input was sufficient to parse the auditory strings into *AXC* words, without any exploitation of the remote dependencies. Furthermore, given that Parser implies only ubiquitous mechanisms of associative learning and memory presumably at play in any situation, it appears highly probable that this information was actually exploited by the participants.

Experiment 1

The above simulations were obviously blind to any prosodic information, such as the 25-ms gap introduced between the *AXC* words in some experiments, and were unable to exploit any idiosyncratic features of the materials, such as the fact that the syllable /pu/ was word beginning and the syllable /ki/ was word ending. Now, these features could also be influential on speech segmentation—for instance, because they may provide information about word boundaries in natural language. The present experiment explores the role of these training-independent factors. Because the question at hand is whether segmentation cues are available in the input before any statistical computation, we removed the nonadjacent dependencies from the material while keeping unchanged most of the other features. To do so, we generated a new set of trisyllabic words. The first and the third syllables of each word, as well as the middle syllables, were those of Peña et al. (2002), but all combinations were allowed. Because there were three possible syllables in the initial, intermediate, and final position, respectively, the whole set comprised 27 different words. If the speech flow is perceived as a succession of discrete units in those conditions, this result must be assigned to the effects of training-

independent factors. Indeed, the distant dependencies were broken, and, as a consequence, the transitional probabilities, whatever the order of the dependency rules that one considers, were exactly the same for any syllable composing the speech stream.

One may reasonably infer from the description of Peña et al.'s (2002) materials that the input provided no prosodic cue to parse the speech flow into the trisyllabic words composing the materials. This conjecture is logical for the experiments including no pauses. Furthermore, generalization of this inference to the experiments in which a 25-ms gap was introduced between the words seems just as straightforward, because Peña et al. repeatedly claimed that the "pauses are subliminal," hence leading to "conditions subjectively very similar" to the no-pause situation (p. 606). The source of Peña et al.'s assertion was an additional experiment reported in Endnote 22. Participants were exposed to two sequences, each lasting 1 min. One sequence included the pauses, and the other did not. Subsequently, participants were asked whether they had heard any silent gaps, and 13 out of the 14 participants responded negatively. In addition, participants responded at chance levels when asked which of the two sequences included pauses.

However, the crucial point is not whether participants detected the pauses but whether they perceived the auditory string including the pauses as a succession of words. It is possible that participants perceived words without being able to indicate what cues guided their perception. Subjective experience with natural languages suggests that the case is far from implausible. For us to test this hypothesis, the questions asked of the participants needed to focus on whether the participants perceived the speech as a succession of discrete words—that is, on the effects of pauses rather than on the pauses themselves.

We ran a first pilot experiment with a group of 33 undergraduate French students, using the statistically flat language described above. Participants were presented with a continuous 2-min speech stream including 25-ms pauses, synthesized in the same way as Peña et al.'s (2002) materials. After the presentation, they were invited to write their responses to four questions in succession on a sheet of paper. The first question was, "Did you hear any silent gaps?" All the participants responded negatively. Therefore, the presence of the 25-ms pauses remained largely undetected, which replicated Peña et al.'s results. However, the responses to the following questions yielded a different picture. The questions were, "Did you perceive the sequence as a continuous speech flow, or as a sequence of words?" "Were the lengths of the words constant or varied?" (to be answered only by those who answered "sequence of words" to the prior question), and, "The words were composed of how many syllables?" (to be answered only by those who answered "constant" to the previous question). Twenty out of our 33 participants perceived the language as a succession of words. Among them, 14 declared that all the words had equal length, and 12 of those 14 participants declared that the words were composed of three syllables. In summary, although no participants perceived pauses as such, 12 out of 33 participants claimed to perceive the speech flow as a succession of trisyllabic words, in a condition in which the remote dependencies introduced in the Peña et al. (2002) materials had been suppressed.

To explore whether the trisyllabic words that some participants claimed to have perceived matched the stimulus words, we ran a second pilot experiment on another group of 21 undergraduate students. The speech stream was generated as in the prior experiment. That is, the experiment involved 27 different trisyllabic words, obtained by the combinations of the syllables used in Peña et al. (2002), and respected the position of each syllable within a word. Participants were asked to write down the words they perceived while listening to the speech stream. Overall, participants gave 280 items. Among them, 262 were trisyllabic, and all trisyllabic items followed the *AXC* pattern. In terms of individual performances, the responses of 15 of the 21 participants contained exclusively *AXC* items, and all but 1 of the remaining participants gave evidence of a strong bias in favor of the *AXC* items, with performance ranging between 67% and 94% of correct responses. These results show that, overall, participants perceived the language including the 25-ms pauses as a sequence of trisyllabic words, with the word boundaries located at the pauses.

However, the pilot experiments described above remain incomplete in at least two respects. First, they did not demonstrate that the segmentation is due to the introduction of the 25-ms gap, because they did not include a no-gap condition. Second, because of group testing, all participants heard exactly the same sequences, with the same word order. The next experiment was devised to make a comparison between conditions with and without pauses. A control group in which the word structure was broken was also tested. All subjects were tested individually to allow individual randomization procedures.

Method

Participants. A total of 72 undergraduate students from the University of Bourgogne in Dijon, France, participated in the experiment in partial fulfillment of a course requirement. All subjects were native French speakers. Participants were randomly assigned to one of the three experimental groups (gap, no gap, and random) until there were 24 participants for each group.

Materials. For the gap and the no-gap groups, the words were generated as explained above. That is, we obtained the words by combining the syllables used in Peña et al. (2002), respecting the position of each syllable within a word. Because there were three first syllables (/pu/, /be/, and /ta/), three last syllables (/ki/, /ga/, and /du/), and three intermediary syllables (/li/, /ra/, and /fo/), the whole set comprised 27 different words. These words were ordered in such a way that no two consecutive words shared any syllable in common (in Peña et al.'s studies, restrictions were framed in terms of word family, which is not possible here. Nevertheless, those restrictions amount to the same as the criterion adopted here). The two groups differed only on the presence (for the gap group) or the absence (for the no-gap group) of the 25-ms silent gap between words. For the additional control group, the syllables composing the words were presented in random order, with the following restrictions: (a) Each sequence of $27 \times 3 = 81$ syllables included the same number of different syllables, and (b) the repetition of any syllable could not occur before 5 different syllables had been presented. We adopted these restrictions to match as closely as possible the materials for the control group and the other two groups on all factors that were not of immediate concern. Within each group, a different order was generated for each participant, but participants from the gap and no-gap groups were yoked in such a way that the same sequence was played for each pair.

Apart from the differences described above, the artificial language was identical to that of Peña et al. (2002). The speech was synthesized through the MBROLA (Multiband Resynthesis Overlap Add) speech synthesizer (see http://tcts.fpms.ac.be/synthesis/; Dutoit, Pagel, Pierret, Bataille, & Van der Vrecken, 1996) with the FR2 diphone database. The mean syllable duration was 232 ms. The resulting wav file was modified using CoolEdit. Progressive fades in and out were applied to the first and last 5 s of each segment to avoid word boundary cues. The speech stream was played through headphones connected to a personal computer using Windows Media Player.

Procedure. Participants were told that they would listen to short samples of an imaginary language. After hearing the first 27-word sequence (for the gap and no-gap groups) or the first 81-syllable sequence (for the random group), they were informed that this language was in fact composed of words. They were told that other, similar samples of the language would be played and that, while listening, they would have to write down on a sheet of paper the words that they perceived. They were asked to avoid engaging in analytic, problem-solving activities. They then listened to four successive samples of the language. Each sample was composed of one instance of each different word for the gap and the no-gap group and of the same pool of syllables in randomized order for the random group. Each block lasted about 20 s, and a 3-s pause separated the blocks. Participants were given a new response sheet for each block. This procedure was aimed at facilitating changes in participants' mode of segmentation throughout the experiment.

Results and Discussion

Participants from the gap, no-gap, and random groups gave, respectively, a mean of 21.00, 17.25, and 15.00 polysyllabic words (SE = 1.34, 0.97, and 0.97, respectively) over the four blocks. The differences were significant, $F(2, 69)$ = 7.48, p = .001, MSE = 29.49, η_p^2 = .178. Orthogonal decomposition revealed that the gap group produced more items than the other two groups, $F(1, 69)$ = 12.90, p < .001, MSE = 29.49, η_p^2 = .157, which did not differ significantly from each other, $F(1, 69)$ = 2.06, p = .16, MSE = 29.49. To assess the other effects independently, in the following analyses we use as the dependent variable a ratio computed over the total number of written words.

Figure 1 shows the proportion of trisyllabic words among the production of the three groups of participants. An analysis of variance (ANOVA) performed on these data revealed a main effect of groups, $F(2, 69)$ = 25.63, p < .0001, MSE = 0.076, η_p^2 = .426. The gap group produced a proportion of trisyllabic items that far exceeded that of the other two groups, $F(1, 69)$ = 45.98, p < .0001, MSE = 0.076, η_p^2 = .400. However, the no-gap group wrote a larger proportion of trisyllabic items than the random group, $F(1, 69)$ = 5.28, p = .024, MSE = 0.076, η_p^2 = .071.

A final analysis concerned the proportion of the words respecting the *AXC* pattern among the trisyllabic words that were written by the participants. Among the trisyllabic words written by the gap group, 96.97% (SE = 0.02) followed the *AXC* pattern. The corresponding value for the no-gap group was 72.35% (SE = 0.06; 1

participant from this group gave no trisyllabic words and was removed from this analysis), a value that was significantly lower than that of the gap group, $F(1, 44)$ = 15.73, p < .001, MSE = 0.044, η_p^2 = .263. This difference was no surprise. However, the performance of the no-gap group was unanticipated. Indeed, given the sequence, a randomly drawn item of three adjacent syllables has one chance out of three of respecting the *AXC* pattern (the other two equally probable patterns are *XCA* and *CAX*). The observed performance for the no-gap group dramatically exceeded this level.

This result suggests that the remarkable segmentation performance of the gap group is not entirely due to the introduction of a 25-ms gap between words. The performance of the no-gap group can hardly be attributed to the statistical properties of the materials, insofar as both adjacent and nonadjacent dependencies were removed from the speech flow. It is likely that participants were sensitive from the start to some prosodic or phonological aspect of the auditory strings.[2] Onnis, Monaghan, Chater, & Richmond (2003) reached a similar conclusion through a set of independent experiments. The possibility that participants exploited a peculiarity of the Peña et al. (2002) materials was pointed out by Seidenberg, MacDonald, and Saffran (2002). These authors noted that all of the first and final syllables of Peña et al.'s words began with a stop consonant, whereas all of the medial syllables began with a continuant consonant. If one labels a stop consonant *S* and a continuant consonant *C*, the participants in fact listened to a sequence *SCSSCSSCSSCS* . . . . This sequence could elicit a rhythmic structure prompting chunking into trisyllabic words, because only this mode of parsing (by steps of three or multiples of three) can generate chunks of identical structure. Thus, this factor could account for the fact that the experimental groups perceived more trisyllabic words than the random group, in which this ternary structure was absent.

But why should such a sequence be chunked as *SCS* units, rather than *CSS* or *SSC*? The beginning and the end of the auditory sequences could not help, given that they were shadowed by a 5-s fade. To address this question, we looked at whether *SCS* patterns are more frequent than *CSS* or *SSC* patterns in French language, on the one hand, and in the production of participants from the random group, on the other hand. Regarding French language, we listed the number of trisyllabic words matching with one of the three patterns of concern in the database LEXIQUE (New, Pallier, Ferrand, & Matos, 2001; see also Peereman & Dufour, 2003). There were 1,380, 1,528, and 2,058 words matching with the *SCS*, *CSS*, and *SSC* patterns, respectively. Therefore, the hypothesis that *SCS* patterns could be predominant in the native language does not hold. Regarding the production of the random group, 63 out of the 121 trisyllabic words written by the participants belong to one of the three relevant structures. There were 23, 21, and 19 words matching the patterns *SCS*, *CSS*, and *SSC*. A student *t* computed

Figure 1. Mean proportion of trisyllabic words over the total number of words for each of the three groups of participants in Experiment 1. Error bars represent standard errors.

---

[2] We do not intend to suggest that the choice of the Peña et al. (2002) material was especially defective. Rather, we believe that such biases are almost unavoidable for a variety of reasons, and this state of affairs emphasizes the need for careful counterbalancing, randomization, or controls in work with artificial languages (see Onnis, Monaghan, Chater, & Richmond, 2003; Perruchet & Reber, 2003; Reber & Perruchet, 2003). This recommendation echoes the claims of Forster (2000) for natural language.
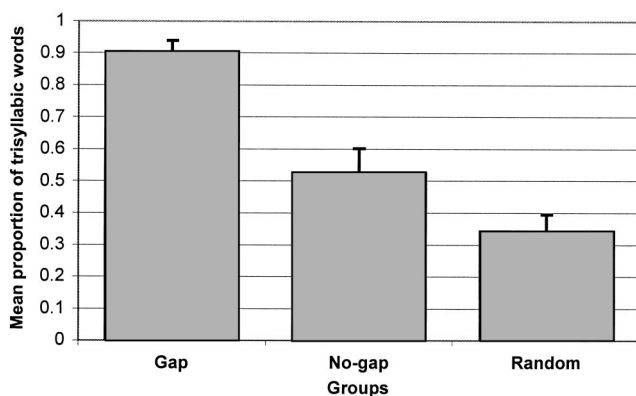
over the 21 participants who gave at least 1 word belonging to one of the relevant structures showed that *SCS* was not produced significantly more often than the other two patterns, $t(20) = 0.57$, $p = .573$. Therefore, although an analysis considering the class of the consonants may have explained why the experimental groups perceived more trisyllabic words than the random group, independent evidence fitting with the observed predominance of *SCS* patterns is lacking.

The material contains many other idiosyncratic features that participants could have exploited. To account for their own results, Onnis, Monaghan, et al. (2003) reported some similarity between the distribution of the initial consonants in Peña et al.'s (2002) syllables and in the French computerized database BRULEX (Content, Mousty, & Radeau, 1990). To explore further their explanation, we conducted a similar analysis with the database LEXIQUE, which contains many more lexical entries than BRULEX (128,972 vs. 35,746, respectively). As indicated in Table 1, we failed to find any confirmatory evidence. As far as there are any differences that can be considered worthy of note, they run against Onnis et al.'s suggestion (although their explanation could be relevant for English).

We also explored the role of vowels and the role of the whole consonant–vowel syllables. In each case, we looked at both the French corpus LEXIQUE and the production of the random group. To make a long story short, the most positive evidence comes from the syllable /pu/, which is one of the three syllables beginning Peña et al.'s (2002) words. Indeed, /pu/ never occurs at the end of French words comprising three or more syllables and occurs in final position for only 4 bisyllabic words. In line with this regularity, participants from the random group wrote 63 polysyllabic words beginning with /pu/, against only 9 polysyllabic words ending with the same syllable. However, similar analyses on other speech segments gave quite mixed and sometimes contradictory results, so that, overall, the relationships between participants' productions and the regularities present in participants' native language appear tenuous at best.

In a final attempt to account for the observed pattern of results, we explored whether they could be explained, in part, by the parameters used to generate the synthetic speech. A potentially relevant feature concerned the phoneme durations. In both Peña et al.'s (2002) and our auditory strings, all phonemes, whether consonant or vowels, were assigned the same duration (116 ms). This is certainly longer than the mean duration of an unvoiced plosive consonant in French (we observed a range from 60 to 120 ms, with a median around 75 ms, in informal analyses of various samples of speech). Given that silent gaps naturally occur before unvoiced plosives in speech, the silent gaps preceding unvoiced consonants turn out to be longer than in natural speech (because the duration of the silence is proportional to the duration of the whole phoneme, at least with the MBROLA synthesizer). These unusually long gaps could have assisted word segmentation in Peña et al.'s language because two of the three unvoiced consonants are in word initial position (/p/ and /t/), whereas only one of them begins a terminal syllable (/k/). The final conclusion of all these exploratory analyses is that the strong perceptual biases evidenced in our Experiment 1 certainly originate from a variety of factors, the respective contributions of which should warrant further studies.

To summarize, we showed, through two simulation studies in which the effect of training-independent factors were prevented, that Peña et al.'s (2002) materials included statistical regularities besides the target *A–C* relationships and that these regularities were sufficient to allow word segmentation. Then we performed an experiment in which the possibility of learning the *AXC* words from the statistical regularities embedded in the materials was removed, to explore the effect of training-independent factors. Participants perceived an overwhelming proportion of the words following the *AXC* pattern (see also Onnis, Monaghan, et al., 2003, for a similar conclusion). If the speech stream was directly perceived as a set of *AXC* words, it follows that the part words used during the test were not even encoded during the familiarization phase. In those conditions, the fact that participants failed to recognize them becomes trivial (certainly the reader would fail to recognize that the sequence /liper/ was present a few lines above, in "directly perceived," without explicitly rehearsing the whole sentence). Overall, these results cast strong doubts on the reliability of the test of remote dependency learning implemented in Peña et al.'s study, which relies on the postulate that the discovery of remote dependencies between *A* and *C* is the only possible means of segmenting the speech stream into *AXC* words.

Experiment 2

The prior computational and experimental data provided evidence that both training-dependent and -independent irrelevant factors could have affected Peña et al.'s (2002) results, because the test they used was contaminated by these factors. However, we did not demonstrate that these factors were actually influential on word learning and, if they were, the extent to which our reanalysis undermines Peña et al.'s conclusions. To address this question, we ran a new experiment with exactly the same training conditions as in one of Peña et al.'s experiments (Experiment 5) and then compared the performance obtained by the participants in the Peña et al. test with their performance in the methodologically sound test described in the introduction. If we obtain a difference between the two measures, this will provide concrete evidence that Peña et al.'s data are unreliable.

Table 1
Percentage of Words Beginning With Each Consonant in Peña et al.'s (2002) Language

Position and phoneme	BRULEX	LEXIQUE
Initial		
/p/	6.67	7.75
/b/	2.20	2.80
/t/	5.00	3.67
Total	13.87	14.21
Medial		
/l/	3.95	1.65
/r/	4.60	10.38
/f/	2.02	3.01
Total	10.57	14.94
Final		
/k/	8.92	8.72
/g/	1.22	1.38
/d/	6.51	9.91
Total	16.65	20.00

Note. BRULEX and LEXIQUE are French computerized databases. The values reported for BRULEX were computed by Onnis, Monaghan, et al., 2003.

Method

Participants. Forty undergraduate students from the same pool as in the prior experiments took part in an individual session, in partial fulfillment of a course requirement. No person participated in more than one experiment of the present series.

Materials. For half of the subjects ($n = 20$), the study material was identical to that used in Peña et al.'s (2002) experiments. Thus, there were three families of words, *pu__ki, be__ga,* and *ta__du,* with three intervening syllables, /li/, /ra/, and /fo/. The test materials were composed of three categories of items (the first two were the items used by Peña et al., 2002): (a) rule words. They belonged to the families used during the familiarization phase, but they were presented with a new intervening syllable. They were *pubeki, putaki, pugaki, beduga, bekiga, bepuga, tagadu, tabedu,* and *takidu.* (b) part words. They straddled word boundaries. They were *kitara, kitafo, gapufo, dubera, likita, lidube, radube, ragapu,* and *fogapu.* (c) scrambled words. They were similar to words (in particular, their intermediary syllables were not heard in the same location during the familiarization phase), but the first and last syllables were shuffled between families. They were *pubega, pukidu, putadu, beduki, begaki, bepudu, tagaki, tabega,* and *takiga.* Note that the intermediary syllables used for the three categories of items were the same and were taken from Peña et al.

For the other half of the subjects ($n = 20$), the material was identical except that /ki/ was replaced by /du/, /du/ was replaced by /ga/, and /ga/ was replaced by /ki/ throughout the study and the test phases. This change was needed because subjects may have a natural tendency to choose, for instance, *pubeki* (a rule word) over *pubega* (a scrambled word) before any experience with the material (Reber & Perruchet, 2003). To circumvent this possibility, Gomez (2002) used two artificial languages, with the words for a language being the scrambled words used for testing the participants trained with the other language, and vice versa. We implemented a similar procedure; however, because of the structure of the material used by Peña et al. (2002), the reversal could not apply to all the items. Although all the items were changed, the study–test reversal concerned only two thirds of the items.

As in the prior experiments, the speech was synthesized through the MBROLA speech synthesizer with the FR2 diphone database. The mean syllable duration was 232 ms. Progressive fades in and out were applied to the first and last 5 s of the familiarization stream. The speech stream was played through headphones connected to a personal computer.

Procedure. Participants were told that they would listen to a 3-min sample of an imaginary language and that they would have to answer questions about the language at the end of the presentation. They were asked to avoid engaging in analytic, problem-solving processes. Each of the nine words was repeated 30 times in the familiarization phase. A silent period of 25 ms was introduced after each word. The words were pseudo-randomly ordered for each subject, except that two words belonging to the same family, and/or sharing the same intermediate syllable, were not adjacent, as in Peña et al. (2002). After familiarization, participants were told that they would be presented with pairs of items and that they would have to judge, for each pair, which item seemed more like a word of the imaginary language. There were 27 pairs of items, with each pair separated by a 500-ms interval. Nine pairs contrasted rule words and part words. This comparison was introduced to replicate Peña et al.'s procedure. Nine pairs contrasted rule words and scrambled words. They provided the correct comparison. Finally, nine pairs contrasted scrambled words and part words. This comparison was aimed at capturing the overall effect of the irrelevant factors that may have influenced the contrast used by Peña et al.

The specific items chosen for each of the 27 pairs were determined randomly for each participant, and for each participant the order of the items within a pair and the order of the pairs in the test sequence were also randomized.

Results and Discussion

The scores obtained in the test were submitted to a 2 (materials) × 3 (type of paired comparisons) ANOVA with repeated measures on the last factor. The main effect of type of comparison was significant, $F(2, 76) = 4.03$, $p = .022$, $MSE = 0.041$, $\eta_{p}^2 = .096$, but there was no main effect of materials and no interaction between materials and type of comparison ($Fs < 1.00$ in both cases). As a consequence, the data were pooled over the two languages.

Results are shown in Figure 2. Rule words were selected over part words for 69.5% ($SE = 3.89$) of the pairs, a score that differs from chance (50%), $t(39) = 5.01$, $p < .0001$. This result closely replicates Peña et al.'s (2002) observation. Peña et al. reported a score of 69.6% with a 10-min familiarization phase (Experiment 3) and a score of 67.1% when exposure was reduced to 2 min. However, participants in the present experiment selected scrambled words over part words on 64.8% ($SE = 3.41$) of the trials, a result that was also above chance, $t(39) = 4.34$, $p < .0001$. In fact, when faced with a pair consisting of a part word and either a scrambled word or a rule word, participants selected the scrambled words, in which the target relationships had been inverted, almost as often as the rule words. Statistically, the difference between the two scores failed to reach significance at the conventional .05 threshold, $F(1, 39) = 2.86$, $p = .098$.

As argued above, the relevant test contrasted rule words and scrambled words. Participants selected the rule word over the scrambled word for 56.7% ($SE = 2.82$) of the pairs, a score that significantly differed from chance, $t(39) = 2.38$, $p = .022$. Thus, this experiment demonstrates the possibility of detecting remote dependencies with Peña et al.'s (2002) materials. This result cannot be due to the particular consonantal pattern of Peña et al.'s (2002) language that was pointed out by Seidenberg et al. (2002). Indeed, both rule words and scrambled words share the same stop/continuant/stop pattern. However, the effect ascribable to the learning of remote dependencies was significantly lower than the effect observed with Peña et al.'s comparison, $F(1, 39) = 6.21$, $p = .017$, $MSE = 0.052$, $\eta_{p}^2 = .137$. Considering the Cohen's effect size conventions (e.g., Cohen, 1988; $d = .20$ is small, $d = .50$ is medium, and $d = .80$ is large), the genuine effect (rule words vs. scrambled words) fell between small and medium (Cohen's $d = .38$), whereas the effect measured with Peña et al.'s test (rule
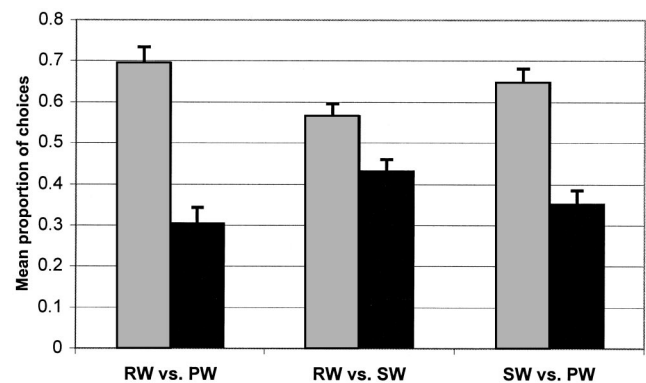

Figure 2. Mean proportion of participants' choices according to the paired comparison in Experiment 2. Error bars represent standard errors. RW = rule words; PW = part words; SW = scrambled words.

words vs. part words) was quite close to the value defining a large effect ($d = .79$). Our third comparison (scrambled words vs. part words) was designed to assess the influence of the irrelevant factors included in Peña et al.'s test. The effect of these factors fell between medium and large ($d = .69$). Thus, we find it interesting that our study allows a roughly additive decomposition of the effect reported by Peña et al., with the smaller (although reliable) part attributable to the genuine discovery of the remote dependency and the larger part to irrelevant factors.

General Discussion

Peña et al. (2002) measured the learning of remote dependencies between *A* and *C* through a test whose validity rests on the assertion that this form of learning is a cause, moreover the exclusive cause, of participants' segmentation of the speech stream into the *AXC* units.

In the first part of our article, we investigated whether this assertion is realistic. We first showed, through two simulation studies, that Peña et al.'s (2002) materials included statistical regularities other than the *A–C* relationships and that those regularities were sufficient for participants to learn the words composing the speech stream. Indeed, Parser, a computational model that is by principle unable to exploit remote dependencies and is blind to any specific properties of the input, discovered the *AXC* words from the input presented to actual participants. Then we designed an experiment (Experiment 1) in which any possibility of learning from dependency relationships was removed but in which all the other aspects of Peña et al.'s materials were preserved. When a 25-ms lag was introduced between the *AXC* words, as in Peña et al.'s Experiments 3 and 5, participants wrote an overwhelming proportion of words respecting the *AXC* pattern, when asked to write what they heard as they heard it. Their performance was statistically better than the performance of a group hearing the same language without the 25-ms lags, as used in Peña et al.'s Experiments 1, 2, and 4. This result runs against the repeated claim of Peña et al. that the gaps do not change the learner's subjective experience (see also Onnis, Monaghan, et al., 2003, for data leading to the same conclusion). Moreover, participants from the no-gap group also perceived many more *AXC* words than could be expected by chance. Thus, both training-dependent and -independent irrelevant factors combined to help segmentation of the speech stream, hence invalidating the use of the participants' speech segmentation abilities as a measure of remote dependency learning.

Experiment 2 was designed to compare the score obtained in the Peña et al. (2002) test with the score given by the correct method of assessing the learning of remote dependencies, in an experiment based on Peña et al.'s Experiment 5. The size of the effect measured by Peña et al.'s test was large according to Cohen's conventions. By contrast, the genuine effect of remote dependencies learning, although significantly above chance, fell between low and medium. The effects due to the factors favoring the segmentation, hence entering into Peña et al.'s score but irrelevant to the learning of distant dependencies, were measured through an independent test, and they were found to be reliably stronger than the effect of interest.

Theoretical Issues

We have now to examine the implications of these results with regard to the longstanding debate between those who advocate the need for assuming algebraic-like computations (e.g., Marcus et al., 1999; Pinker, 1997) and the proponents of statistical/distributional approaches (e.g., Redington & Chater, 1998; Seidenberg & MacDonald, 1999).

A first, straightforward conclusion is that the Peña et al. (2002) study lends no support to the existence of algebraic-like computations. If one replaces what Peña et al. intended to measure by what they actually measured, their final line of argumentation stands as follows. They (a) observed that word segmentation is not possible when *AXC* words are displayed as a continuous speech stream, (b) introduced an explicit cue for segmentation between the *AXC* words, (c) observed that this cue indeed helps segmentation, (d) inferred from this observation that it is possible to detect the remote dependencies between *A* and *C* when segmentation cues are displayed, and (e) attributed this achievement to the fact that the segmentation cues made the stream more similar to natural language, hence triggering language-specific algebraic-like computations. It is clear that Points d and e do not logically follow from Points a–c. What Peña et al. showed is that adding segmentation cues helps segmentation, and that is all.

However, we do not claim that this reappraisal rules out the hypothesis that algebraic mechanisms are actually involved in the procedures devised to reveal learning of nonadjacent dependencies. It is impossible to demonstrate that there is no such thing as a rule abstractor device, because demonstrating nonexistence lies beyond any empirical approach. The only possible endeavor consists of demonstrating that such a device is useless, because the available empirical evidence is compatible with another, more parsimonious approach. Now, even though the effect was far less impressive than reported by Peña et al. (2002),[3] we found reliable evidence for the possibility of learning nonadjacent dependencies between syllables. Accounting for this effect within an associative view of mind does not seem to be a trivial task, in view of the fact that Parser (Perruchet & Vinter, 1998), which is rooted in associative learning principles, is unable to do so. Recall, indeed, that this incapability was exploited to demonstrate that Peña et al.'s language included other statistical cues for segmentation.

We argue, nevertheless, that, somewhat paradoxically, the available pattern of data fits nicely with the predictions of the general theory of associative learning on which Parser relies. Parser is based on a view of associative learning and memory in which associations are closely dependent on attentional processes. This principle is fundamental in the literature on conditioning in animals (Mackintosh, 1975) and has become largely accepted in the context of the studies on implicit learning (e.g., Perruchet & Vinter, 2002; Whittlesea & Dorken, 1993). More generally, many authors, using different terminologies, have proposed a view com-

---

[3] It is worth noting that researchers committed to a rule-based framework often tend to overestimate the human ability to learn syntactically complex linguistic structures. For instance, Fitch and Hauser (2004) claimed that humans easily master a center-embedded artificial grammar, an achievement that also requires the detection of nonadjacent dependencies. Perruchet and Rey (in press) showed that Fitch and Hauser's procedure had major drawbacks and obtained no evidence for the learning of center-embedded structures when a methodologically sound test was used.

patible with the claim that associative learning is an automatic process that associates all the components that are present in the attentional focus at a given point (Frensch & Miner, 1994; Logan & Etherton, 1994; Perruchet & Vinter, 2002; Stadler, 1995; Treisman & Gelade, 1980). This position, however, begs the question of which components of the environment are concurrently held in the attentional focus. The option implemented in Parser is that attention is focused primarily on contiguous events. This is certainly reasonable when the language is perceived as a continuous, unbroken sequence of syllables. Indeed, there is overwhelming evidence that people segment such sequential material into chunks of a few successive elements. This mode of segmentation is consistent with the reported difficulty to learn nonadjacent dependencies in those conditions (Newport & Aslin, 2004[4]; Peña et al., 2002, Experiments 2 and 4).

However, there is no reason to think that attention necessarily focuses on contiguous events, whatever the sensory input. We showed that participants in Peña et al.'s (2002) experiments perceived not a continuous speech stream but instead a succession of trisyllabic words. Now, both experimental evidence and everyday experience suggest that the start and the end of a sequence capture more attention than the intermediary events. Thus, it is likely that when the auditory stream is perceived as a succession of artificial words, participants pay more attention to the first and last syllables than to the middle one and then encode those syllables as well as the relevant positional information. This prompts the formation of *AXC* units, where *A* and *C* are specific syllables and *X* stands for unnoticed events.

This interpretation of nonadjacent dependencies learning, centered around the role of attentional mechanisms, echoes that of Gomez (2002) and finds support in her results. Gomez showed that, in a situation in which the successive *AXC* units were perceptually distinct, the degree to which the *A–C* relationships were learned depended on the variability of the middle element (*X*). More precisely, participants were presented with 2, 4, 8, or 24 different *X* elements, and it was observed that learning increased markedly under conditions of greatest variability in both adults and infants (see also Onnis, Christiansen, Chater, & Gomez, 2003). Gomez argued that the high variability of the intermediary element led participants to focus attention on the nonadjacent elements, because they appeared to be the more stable features in the situation.

Peña et al. (2002) argued that the possibility of discovering remote dependencies after only 2 min of exposure to the language indicated that this form of learning was not due to the extraction of statistical patterns, because statistical computation would have required a larger amount of exposure. This finding calls for empirical confirmation with a methodologically sound test. In our study, the duration of exposure was 3 min, and we found an effect of quite moderate size, which reached significance because of our large sample size.[5] However, for the sake of argument, let us assume that 2 min would have been sufficient. The assertion that associative learning proceeds slowly does not stand up to empirical observations. For example, some associative forms of learning have been shown to develop over one trial or so (e.g., Fanselow, 1990). Closer to the domain of concern here, Saffran, Aslin, and Newport (1996) observed that babies were able to segment an artificial language presented as a continuous speech flow after only 2 min of exposure. Now, this phenomenon is commonly attributed to statistical mechanisms, even by Peña et al. Overall, the learning

of the *A–C* relationships proceeded at a rate that, roughly speaking, is quite compatible with a statistical or distributional approach.

Conclusion

The detection of remote dependencies represents an apparent challenge for a framework based on associative or statistical mechanisms. Indeed, most of the literature on associative learning has focused on adjacent dependencies, whereas there is evidence, especially in the language area, that higher cognitive activities also exploit the existence of nonadjacent dependencies in the material they deal with. Peña et al. (2002) argued that learning nonadjacent dependencies requires algebraic-like computations, on the basis that such learning is possible only if subliminal segmentation cues make the signal more similar to natural language. When using Peña et al.'s methodology, we found the same empirical results. However, when methodological drawbacks were controlled for, the pattern of results was different and fairly consistent with the action of ubiquitous mechanisms of associative memory and learning. The current attempts to explore the learning of nonadjacent dependencies within a statistical approach (Gomez, 2002; Newport & Aslin, 2004) and to simulate its key aspects with connectionist models (Christiansen & Chater, 1999) appear to be much more promising avenues of research than the rule-based framework suggested by Peña et al. (2002).

---

[4] Newport and Aslin (2004) failed to obtain evidence of learning between nonadjacent *syllables* when the speech flow was not segmented. They also claimed that, by contrast, such learning is possible between nonadjacent *segments* (patterns among consonants skipping over vowels, or patterns among vowels skipping over consonants). The differences in learnability of the distant relationships involving syllables and subsyllabic units is potentially interesting. However, it should be noted that, in their experiments manipulating segments (Experiments 2 and 3), Newport and Aslin compared words with part words straddling over word boundaries, as did Peña et al. (2002). As shown in our Experiment 2, this test provides a striking overestimation of learning. In addition, Newport and Aslin used the same intermediary segment during the familiarization phase and during the test phase, hence not testing for generalization. Overall, the conclusion that "non-adjacent segment regularities . . . are extremely easy to acquire" (Newport & Aslin, 2004, p. 154) needs to be confirmed.

[5] The sample size was 40 participants. When tested on the first 14 participants—that is, with the sample size used in Peña et al. (2002)—the effect did not reach significance, $t(13) = 1.47$, $p = .164$.

References

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

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23,* 157–205.

Cleeremans, A. (1993). *Mechanisms of implicit learning: A connectionist model of sequence processing.* Cambridge, MA: MIT Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 113–163). Hillsdale, NJ: Erlbaum.

Content, A., Mousty, P., & Radeau, M. (1990). Brulex: Une base de données lexicales informatisée pour le français écrit et parlé [Brulex: A

lexical computerized database for written and spoken French]. *Année Psychologique, 90,* 551–566.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. In H. T. Bunell & W. Isardi (Eds.), *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 1393–1396). Wilmington, DE: Alfred I. duPont Institute.

Fanselow, M. S. (1990). Factors governing one-trial contextual conditioning. *Animal Learning and Behavior, 18,* 264–270.

Fitch, W. T., & Hauser, M. D. (2004, January 16). Computational constraints on syntactic processing in a nonhuman primate. *Science, 303,* 377–380.

Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition, 28,* 1109–1115.

Frensch, P. A., & Miner, C. S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory & Cognition, 22,* 95–110.

Gomez, R. (2002) Variability and detection of invariant structure. *Psychological Science, 13,* 431–436.

Logan, G. D., & Etherton, J. L. (1994). What is learned during automatization? The role of attention in constructing an instance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1022–1050.

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

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999, January 1). Rule learning by seven-month-old infants. *Science, 283,* 77–80.

McClelland, J., & Rumelhart, D. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2). Cambridge, MA: MIT Press.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur Internet: LEXIQUE [A lexical database on the Internet about contemporary French: LEXIQUE]. *L'Année Psychologique, 101,* 447–462.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48,* 127–162.

Onnis, L., Christiansen, M. H., Chater, N., & Gomez, R. (2003). Reduction of uncertainty in human sequential learning: Preliminary evidence from artificial grammar learning. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th annual meeting of the Cognitive Science Society, 2003* (pp. 886–891). Boston: Cognitive Science Society.

Onnis, L., Monaghan, P., Chater, N., & Richmond, K. (2003). *Phonology impacts segmentation and generalisation in speech processing.* Manuscript submitted for publication.

Peereman, R., & Dufour, S. (2003). Un correctif aux notations phonétiques de la base de données Lexique [A corrective to the phonemic notations of the Lexique database]. *L'Année Psychologique, 103,* 103–108.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002, October 18). Signal-driven computations in speech processing. *Science, 298,* 604–607.

Perruchet, P., & Peereman, R. (2004). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics, 17,* 97–119.

Perruchet, P., & Reber, R. (2003). Why untrained control groups provide invalid baselines: A reply to Dienes and Altmann. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56A,* 125–130.

Perruchet, P., & Rey, A. (in press). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin and Review.*

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language, 39,* 246–263.

Perruchet, P., & Vinter, A. (2002). The self-organized consciousness. *Behavioral and Brain Sciences, 25,* 297–388.

Perruchet, P., Vinter, A., Pacteau, C., & Gallego, J. (2002). The formation of structurally relevant units in artificial grammar learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55A,* 485–503.

Pinker, S. (1997). *How the mind works.* New York: Penguin Books.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences, 6,* 456–463.

Ramscar, M. (2002). The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology, 45,* 45–94.

Reber, R., & Perruchet, P. (2003). The use of control groups in artificial grammar learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 56A,* 97–116.

Redington, M., & Chater, N. (1998). Connectionist and statistical approaches to language acquisition: A distributional perspective. *Language and Cognitive Processes, 13,* 129–191.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22,* 425–469.

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 585–594.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December 13). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

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

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition, 69,* 105–134.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23,* 569–588.

Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002, October 18). Does grammar start where statistics stop? *Science, 298,* 553–554.

Stadler, M. A. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 674–685.

Stadler, M., & Frensch, P. (1998). *Handbook of implicit learning.* Thousand Oaks, CA: Sage.

Stanback, M. L. (1992). Syllable and rhyme patterns for teaching reading: Analysis of a frequency-based vocabulary of 17,602 words. *Annals of Dyslexia, 42,* 196–221.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Whittlesea, B. W. A், & Dorken, M. D. (1993). Incidentally, things in general are incidentally determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General, 122,* 227–248.