Available online at www.sciencedirect.com
SCIENCE DIRECT®
Cognitive Psychology 48 (2004) 127–162
Cognitive Psychology
www.elsevier.com/locate/cogpsych

Learning at a distance
I. Statistical learning of non-adjacent dependencies

Elissa L. Newport* and Richard N. Aslin

Department of Brain and Cognitive Sciences, Meliora Hall – River Campus, University of Rochester, Rochester, NY 14627, USA

Accepted 1 July 2003

Abstract

In earlier work we have shown that adults, young children, and infants are capable of computing transitional probabilities among adjacent syllables in rapidly presented streams of speech, and of using these statistics to group adjacent syllables into word-like units. In the present experiments we ask whether adult learners are also capable of such computations when the only available patterns occur in non-adjacent elements. In the first experiment, we present streams of speech in which precisely the same kinds of syllable regularities occur as in our previous studies, except that the patterned relations among syllables occur between non-adjacent syllables (with an intervening syllable that is unrelated). Under these circumstances we do not obtain our previous results: learners are quite poor at acquiring regular relations among non-adjacent syllables, even when the patterns are objectively quite simple. In subsequent experiments we show that learners are, in contrast, quite capable of acquiring patterned relations among non-adjacent segments—both non-adjacent consonants (with an intervening vocalic segment that is unrelated) and non-adjacent vowels (with an intervening consonantal segment that is unrelated). Finally, we discuss why human learners display these strong differences in learning differing types of non-adjacent regularities, and we conclude by suggesting that these contrasts in learnability may account for why human languages display non-adjacent regularities of one type much more widely than non-adjacent regularities of the other type.

* Corresponding author.
E-mail address: newport@bcs.rochester.edu (E.L. Newport).

0010-0285/$ - see front matter
doi:10.1016/S0010-0285(03)00128-2

1. Introduction

A question of long-standing interest concerns the mechanisms by which human learners acquire their native language. We know, from numerous empirical studies and theoretical discussions, that this process requires contributions from both nature and nurture—that is, from both the linguistic environment to which learners are exposed and some innate predispositions of human learners to process and learn temporally organized patterns in particular ways (see Chomsky, 1965; Gleitman & Newport, 1995; Marcus, 2001; Pinker, 1994; Seidenberg, 1997; for discussion). However, little is known about the precise processes by which this learning occurs or the mechanisms responsible for its rapidity and success.

In recent work we have shown that adults, young children, and infants are capable of computing transitional probabilities[1] among adjacent syllables in rapidly presented streams of speech, and of using these statistics to group syllables into word-like units (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996a; Saffran, Newport, & Aslin, 1996b; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). We believe this statistical learning mechanism may play an important role in various aspects of language acquisition—at minimum in the process of word segmentation, but also potentially in the acquisition of syntax and morphology as well (Mintz, Newport, & Bever, 2002; Morgan, Meier, & Newport, 1987; Newport & Aslin, 2000; Saffran, 2001, 2002). However, the extent of the capabilities of this statistical learning mechanism, and the levels and types of language patterns that may be acquired with the help of such a computational device, are still unknown. In the present paper we take an important step beyond our earlier results, asking whether learners are capable of computing not only adjacent sound regularities, but also regularities among sounds that are not adjacent to one another, and if so, what types of non-adjacent regularities they can easily acquire.

As noted above, our first work focused on asking whether learners could acquire statistical regularities among immediately adjacent syllables. Indeed, most words in natural languages are comprised of consistent sound patterns among adjacent syllables, and the transitional probability relations we examined in our miniature language studies were similar to those exhibited in real human languages (Harris, 1955). But natural languages exhibit other types of regularities as well, including certain types of non-adjacent patterns (Chomsky, 1957). Any mechanism used broadly in language acquisition must therefore, in some way, be capable of learning non-adjacent regularities (Chomsky, 1957; Miller & Chomsky, 1963).

What types of non-adjacent regularities do natural languages include? In many languages, words contain regular patterns among syllables or phonemic segments

[1] More technically, we have shown that learners compute a conditionalized statistic which tracks the consistency with which elements occur together and in a particular order, baselined against individual element frequency. *Transitional probability* is a particular type of temporally ordered conditional probability, first used for psycholinguistic materials by Miller and Selfridge (1950). But our findings are also compatible with the claim that learners might be computing another closely related statistic, such as *mutual information* or *conditional entropy*.

that are not immediately adjacent. For example, in Tagalog, some words may receive infixes: sounds inserted within the word stem to mark a specific tense or aspect, and in Semitic languages, words may be built from a consonant pattern, such as *k-t-b*, with varying vowel patterns inserted between the consonants to signal time or number. Similarly, syntactic structure may involve dependencies between words that are quite distant from one another: sentence subjects that agree with verbs many words away, or wh-question words that replace noun phrases much later in the sentence. However, a central finding of modern linguistics has been that such non-adjacent relations are quite selective and display limits that are universal to languages of the world; a main enterprise, of theoretical linguistics of all flavors, has been to capture these limitations in a set of principles or universal constraints (Chomsky, 1965, 1981, 1995).

How might a learning mechanism—and in particular, a statistical learning mechanism—operate with regard to non-adjacent dependencies? An important problem for this type of computational mechanism (as for any learning device) concerns how to limit its operations, so that the patterns of language are correctly acquired, but without an unmanageable explosion in the number of computations that must be performed to do the learning (Chomsky, 1965, 1981; Wexler & Culicover, 1980). In order to acquire even the simplest adjacent patterns that we have studied in 4-word, 2-min experiments with infants, learners must be performing the running computation of 20 different transitional probabilities, each over 45 occurrences of the component syllables and 15–45 occurrences of syllable pairings.[2] A learning mechanism additionally capable of computing and acquiring non-adjacent dependencies, while necessary for language learning, opens a computational Pandora's box: In order to find consistent non-adjacent regularities, such a device might have to keep track of the probabilities relating all the syllables one away, two away, three away, etc. If such a device were to keep track of regularities among many types of elements—syllables, features, phonemic segments, and the like—this problem grows exponentially. But, as noted, non-adjacent regularities in natural languages take only certain forms. The problem is finding just these forms and not becoming overwhelmed by the other possibilities.

There are several possible ways of thinking about solutions to this problem. One possibility is that the statistical learning mechanism we have discovered is, in fact, a simple and low-level mechanism, limited to quick calculations among adjacent sound units. If this were the case, it would have to feed its results to another mechanism—perhaps a language acquisition device that is built to expect the properties exhibited by natural languages—in order to acquire the full range of constructions of human languages.

[2] These figures are the number of computations a learner would have to perform to acquire our 'baby' languages in their entirety. At minimum, 8 of the 20 transitional probabilities are tested in our test items and would have to be computed in order to succeed in discriminating words from partwords. In our more complex 'adult' languages with 6 trisyllabic words (cf. Saffran et al., 1996b), the number of transitional probabilities among distinct syllables required for learning is greater; and, of course, it must be *enormously* greater for acquiring real languages.

A second possibility is that the statistical learning mechanism we have discovered might itself be capable of a broader range of computations, among both adjacent and non-adjacent elements. But if so, what kinds of non-adjacent relations is it capable of acquiring? What might be the limits on such a learning device? Is it a very broad computational mechanism, capable of computing many patterns, both those that natural languages exhibit and also those that natural languages do not exhibit? If so, the constraints on patterns that appear in natural languages would have to be provided, during learning, by another source (e.g., a substantive language acquisition device, or a constraint on on-line processing). Alternatively, the particular computations this device can perform and the patterns that natural languages exhibit could be sharply similar and matching in their selectivities. If the latter, this would suggest that some of the constraints on natural language structure might arise from constraints on the computational abilities this mechanism exhibits.

In the present paper we address this question through a series of empirical studies of the learning of non-adjacent regularities. We begin with patterns that are, as much as possible, identical to those we have previously studied, except that they incorporate non-adjacent, rather than adjacent, regularities. As we will see, however, human learning of non-adjacent regularities appears to be extremely selective, even in our laboratory studies. Our studies therefore move on to examine those types of non-adjacent patterns that learners do and do not readily acquire. As we will show, the findings we obtain across these studies match remarkably well with the types of patterns natural languages do and do not commonly exhibit. In a companion paper, we examine this type of learning in a different primate species—Cotton-top tamarin monkeys—to see where these selectivities might be shared or specific to our species. Taken together, these papers begin to shed some light on how statistical learning mechanisms and universals of language might interact.

2. Experiment 1: Non-adjacent syllables

In our previous studies (Aslin et al., 1998; Saffran et al., 1996a, 1996b), subjects readily learned words comprised of consistent sequences of adjacent syllables, discriminating them from non-occurring sequences of the same syllables, and also from sequences of the same syllables that had occurred with less consistency. These results demonstrate that human learners can acquire syllable groupings by computing, online and very rapidly, a set of statistics concerning how adjacent syllables occur in a novel stream of speech. In the present study, we ask whether learners can also demonstrate these same abilities when the groupings are instantiated in statistical regularities only among *non*-adjacent syllables.

In the easiest of our previous languages built on adjacent regularities (Saffran et al., 1996a), 4 tri-syllabic words were built by having unique syllables in each word and words following each other at random (excluding immediate repeats). In this design, the transitional probabilities between syllables within a word were 1.0; the transitional probabilities at word boundaries were .33. This type of language can be learned by 8-month-old infants in only 2 min of exposure (Saffran et al., 1996a).

In the hardest of our previous languages (Saffran et al., 1996b), 6 tri-syllabic words were built using the same 12 syllables, some in only one word but some appearing in more than one word. In this design, the transitional probabilities between syllables within a word were lower than 1.0 but still higher than those at word boundaries. Adult subjects can learn this type of language in 21 min (Saffran et al., 1996b). In order to build a language using non-adjacent regularities, but with approximately this same level of overall complexity, we began by creating a language in which there were 5 regular word frames—sets of syllable pairs that co-occurred with a transitional probability of 1.0. However, we made these syllable pairs non-adjacent, inserting one of 4 different syllables in the middle. The same 4 middle syllables could occur in the middle of all 5 non-adjacent word frames. The words in this language thus all followed a 1–X–3 pattern, with 5 sets of 1–3 instantiations and 4 Xs, for a total of 20 different words in the language.

Given a stream of words, randomly ordered (excluding immediate repeats), following this pattern, there is no strong grouping of syllables if only adjacent syllable relations are computed: the adjacent transitional probabilities in such a stream vary from .25 (at the transition from syllable 1 to syllable X, and also at the transition from syllable 3 to syllable 1 of the next word) to .20 (at the transition from syllable X to syllable 3), with no extremely high or extremely low transitions at any point. However, there is a strong grouping of syllables if learners are able to compute non-adjacent syllable relations: among syllables 1 and 3 (one syllable away from another), the transitional probability is 1.0, whereas the transitional probability among other syllables one away (syllable X and syllable 1 of the next word, or syllable 3 and syllable X of the next word) is only .20 or .25. Exposing subjects to such a stream can thus permit us to ask whether learners are readily able to compute such non-adjacent statistics.

2.1. Method

2.1.1. Subjects

Twenty-four undergraduates were recruited from Brain and Cognitive Sciences classes at the University of Rochester. Subjects were paid $6 for their participation. All subjects were monolingual English speakers, and none had previously been in a statistical learning experiment. All subjects were exposed to a single language (see below) and then tested using one of two test forms, differing only in the randomized order of the test items (13 subjects in order 1 and 11 in order 2).

2.1.2. Stimulus materials

The language consisted of 20 trisyllabic words in which the first and third syllables were perfectly predictable and the second syllable was less predictable. Table 1 shows the design and details of this language.[3] The words were formed from eight

[3] Notation in the tables and text for describing the sounds used in our materials is in the International Phonetic Alphabet (IPA).

consonants (b, d, p, t, g, k, r, l) and five vowels (a, i, u, e, o) which, when combined, rendered an inventory of 14 consonant–vowel (CV) syllables. Ten of these syllables served as the first and last syllables in 5 trisyllabic word-frames (ba_te, gu_do, pi_ra, ke_du, lo_ki). The middle syllable in these five word-frames was filled with each of the four remaining syllables (di, ku, to, pa). The top portion of Table 1 provides a schematic of one word-frame, while the bottom portion lists the actual words. A continuous stream of speech comprised of these words was constructed by forming 6 blocks, each consisting of a different random ordering of the 20 words. These blocks were concatenated into a text in repeated random orders, with the stipulation that the same word, and the same word-frame, never occurred twice in a row. All word boundaries were removed from the text, which was then read by the MacInTalk speech synthesizer, using the text-to-speech application Speaker, running on a Macintosh Quadra 700 computer. Because the synthesizer was not informed of word boundaries, it did not produce any acoustic word boundary cues and produced equivalent levels of coarticulation between all syllables. The speech stream contained no pauses and was produced by a synthetic female voice (Victoria) in monotone, with a speaking rate of 216 syllables per minute. The output of the synthesizer

Table 1
Design of non-adjacent syllable language used in Experiment 1

CV_1 $[CV_{2a}]$ CV_3
$[CV_{2b}]$
$[CV_{2c}]$
$[CV_{2d}]$

1st–3rd Syllable word-frames	2nd Syllables	Words
ba_te	di	ba di te
		ba ku te
gu_do	ku	ba to te
		ba pa te
pi_ra	to	
		gu di do
ke_du	pa	gu ku do
		gu to do
lo_ki		gu pa do
		pi di ra
		pi ku ra
		pi to ra
		pi pa ra
		ke di du
		ke ku du
		ke to du
		ke pa du
		lo di ki
		lo ku ki
		lo to ki
		lo pa ki

was recorded to audiocassette tape for later playback to the subjects. The tape consisted of a 7-min stream of speech which was repeated three times for a total of 21 min of exposure.

This continuous stream of speech created by the random ordering of the 20 words (excluding immediate repetition of any word-frame) resulted in a pattern in which the only available information for extracting words was the greater statistical regularity of non-adjacent syllable sequences within words than of syllable sequences that spanned a word boundary. This grouping could not be learned by computing transitional probabilities only among *adjacent* syllables: the adjacent transitional probabilities within words consisted of a transitional probability of .25 between the first and second syllables within a word, .20 between the second and third syllables within a word, and .25 between the last syllable of a word and the first syllable of the following word. In contrast, the *non-adjacent* statistical structure was a transitional probability of 1.0 between the first and third syllable within a word, as compared with a transitional probability of .20 between the second syllable of one word and the first syllable of the next word and .25 between the third syllable of one word and the second syllable of the next word. Thus, the predictability of adjacent syllables throughout the entire speech stream was low (TP = .20–.25), and the predictability of non-adjacent syllables that spanned a word boundary was also low (TP = .20–.25), whereas the predictability of non-adjacent syllables *within* a word was high (TP = 1.0).

Learning of the statistical coherence of the non-adjacent syllables within words was tested by asking whether subjects could discriminate the *words* from *partwords*. Test items consisted of one of five words (bakute, gupado, pitora, kedidu, lopaki) selected from the inventory of 20 words, paired with one of five partwords (bakudo, gupara, pitodu, kediki, lopate). The partwords were formed by altering the third syllable of each of the five test words such that it came from a different word-frame than the first syllable. As a result, all *adjacent* statistics were matched in the test items, but the *non-adjacent* statistics between the first and third syllables were high in the test words (TP = 1.0) and low in the partwords (TP = .00). Each of the test words and partwords was generated separately by the MacInTalk speech synthesizer, in the same way as described for the streams above, except that each was generated in isolation. This produced a falling intonation on the final syllable of each item, making each (words and partwords) sound like a word spoken in isolation.

A test item consisted of a test word paired with a partword, in counterbalanced order, and was recorded to audiocassette tape with a 1 s silent interval between each word/partword that formed a test pair and a 5 s silent interval between each test pair. The test was a 2-alternative forced choice task, consisting of a randomized pairing of each of the five test words and each of the five partwords (with position in each pair counterbalanced), for a total of 25 test items.

2.1.3. Apparatus

The experiment was conducted in an IAC sound-attenuated booth. The recorded speech stream was presented to subjects using a JVC audiocassette player and an Infinity speaker powered by a Sony amplifier. Test stimuli were presented using the same equipment, and subjects indicated their judgments on a pre-printed answer sheet.

2.1.4. Procedure

Subjects were instructed to listen to a "nonsense" language. They were told that the language contained words, but no meanings or grammar. They were informed that their task was to figure out where the words began and ended. Subjects were given no information about the length or structure of the words or how many words the language contained. They were informed that the listening phase of the experiment consisted of three short segments, followed by a test of their knowledge of the words in the language. Subjects were given a short break after each of the first two 7-min listening segments. All subjects were run individually.

After a total of 21 min of listening, subjects received a two-alternative forced-choice test. For each test trial, subjects heard 2 trisyllabic strings, separated by 1 s of silence. One of these strings was a word from the nonsense language, while the other was a partword, in counterbalanced order. Subjects were asked to indicate which of the two strings sounded more like a word from the language by circling either the "1" or "2" on the answer sheet. The test items were constructed by pairing the 5 test words from the language with 5 trisyllabic foils (the partwords). Each word was paired exhaustively with each foil, rendering 25 test items. There was a 5-s interval following each test trial to allow the subject to record a response on the answer sheet prior to the onset of the next item. Two different randomized orders of presentation of the test items were used (counterbalanced across subjects).

2.2. Results

Fig. 1 shows the results on the 2AFC test, which asked subjects to choose between words of the language (tri-syllabic sequences following the 1–X–3 pattern, which exhibit non-adjacent transitional probabilities of 1.0 in the exposure stream) and partwords (tri-syllabic sequences following a 1–X–4 pattern, in which the adjacent pairs of syllables had occurred in the exposure stream, but in which the 1–4 non-adjacent sequence does not occur). The figure shows the results separately for the 2 test orders. There was no significant effect of test order ($t(22) = .08$, $p = .93$, *ns*). Most important, there was no evidence of learning on either test order, or on the results pooled across test orders. Performance on the 2AFC test did not exceed chance (test order 1: mean = 11.92 out of 25, or 47.7% correct, $t(12) = -.70$, $p = .50$, *ns*; test order 2: mean = 11.82 out of 25, or 47.3% correct, $t(10) = -.71$, $p = .49$, *ns*; overall: mean = 11.88 out of 25, or 47.5% correct, $t(23) = -.20$, $p = .84$, *ns*).

2.3. Discussion

This result clearly shows that subjects did not succeed in learning the non-adjacent regularity in this stream, despite the fact that it was a highly consistent regularity (non-adjacent transitional probability of 1.0) recurring across each of only 5 word frames more than 75 times apiece in the 21-min exposure.

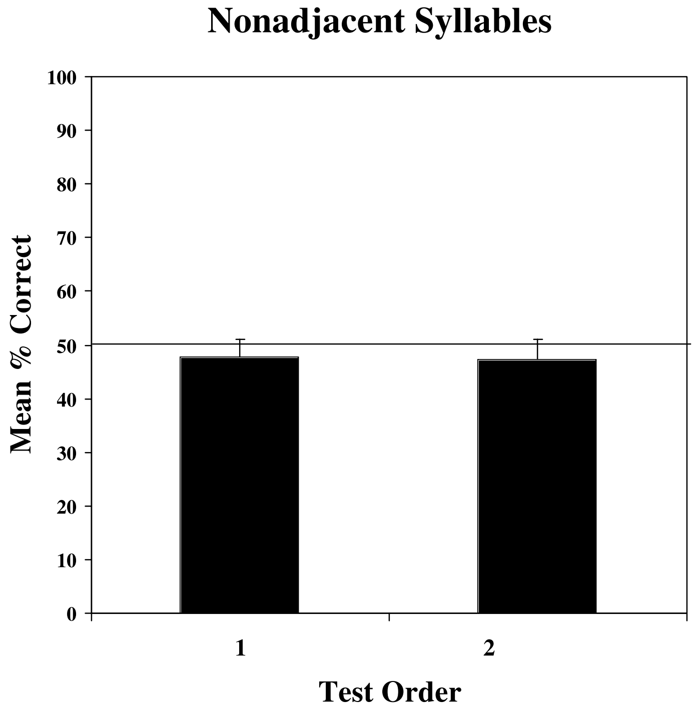

Fig. 1. Mean percent correct on the non-adjacent syllable language of Experiment 1, for two tests differing in item order.

Because the statistics to be learned in this experiment were so similar to those of our previous experiments, except for the non-adjacent feature of the design, this result was at first somewhat surprising. However, there were at least two reasons for this possible failure to learn, quite different from one another in their import. First, subjects could be failing to learn the non-adjacent regularity because the size of the language was too great for learning to occur in only 21 min: 5 word frames and 4 middle syllables formed 20 words, more than we had previously asked participants to learn in a brief experiment. Alternatively, this could be a true difficulty in learning non-adjacent, as compared to adjacent, regularities. To determine which of these accounts was correct, we proceeded to conduct a series of experiments, gradually making the languages smaller and smaller and varying a number of other details of the presentation, to see whether we would find learning of the same non-adjacent pattern when other aspects of the experimental procedure were optimized.

3. Further explorations of these negative findings

Over a series of eight different experiments, involving a total of 51 subjects, we manipulated a number of variables to see whether we could demonstrate successful learning of non-adjacent syllable regularities. First, we increased the length of exposure subjects were given to the language (running some subjects for 2 sessions, across 2 consecutive days, rather than one). We also tried an implicit rather than explicit learning procedure (as in Saffran et al., 1997), since some miniature language learning literature has suggested that implicit learning might produce superior results for complex patterns (Reber, 1976). We ran some subjects on an easier type of test item, asking them to choose between words versus nonwords (the same syllables in an order that never occurred in the exposure stream) rather than words versus partwords. We ran one unfortunate subject for 10 successive days, testing his learning at days 5 and 10. We experimented with different sound choices, to see whether we had accidentally chosen a particularly difficult set of sounds for our initial run on this type of language. And we made the language more and more simple, while preserving the crucial non-adjacent syllable regularity, by moving from a language with 5 word-frames and 4 middle syllables, to one with 4 word-frames and 3 middle syllables, and finally to one with 3 word-frames and 2 middle syllables.

To our initial surprise, none of these variations produced consistent learning (see Fig. 2). Indeed, none of them produced *any* evidence of learning at all, except in one run testing words versus non-words—a discrimination that can be performed without in fact acquiring the non-adjacent syllable pattern. Even the single subject who was run for 10 days scored only 48% correct (12 correct out of 25) on the word versus partword task.

These results demonstrate that learners have some striking difficulty acquiring a very simple pattern of non-adjacent syllable regularities—at least within a paradigm in which many statistical computations, across a number of syllables and syllable positions within a continuous stream, would have to be conducted to find the specific simple and recurring non-adjacent regularity that comprises our language.

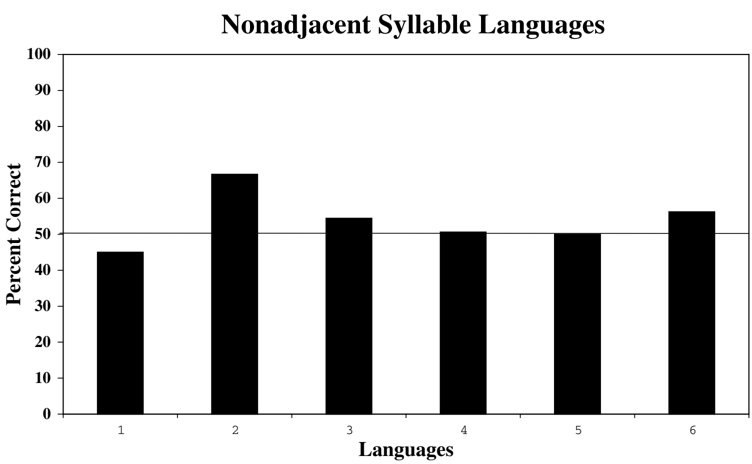

Fig. 2. Mean percent correct on several non-adjacent syllable languages varying in size, learning mode, and type of test item.

We do know that human learners are capable of acquiring patterns between non-adjacent syllables under certain circumstances. For example, in the Marcus et al. paradigm (Marcus, Vijayan, Bandi Rao, & Vishton, 1999), 7-month-old infants can learn an ABA pattern and discriminate it from an ABB pattern. Similarly, in Gomez (2002), both adults and infants can learn an AXB pattern under some conditions, though not under others (see Section 7 for further discussion of this contrast). However, in both of these experimental paradigms, the exposure materials consist of a series of separate 3-syllable sequences, with lengthy pauses before and after each 3-syllable string. The learner in such a paradigm must extract the relationship between non-adjacent syllables when the stream is already parsed, when the total number of syllables in a particular sequence that might exhibit a patterned relation is extremely limited, and when there are no other non-adjacent syllables whose regular relations are in question. In contrast, in our own paradigm, a long and continuous stream of many syllables is presented to the learner; in order to find the particular non-adjacent syllable relation that is highly patterned, the learner would have to conduct many statistical computations, across a number of syllables and syllable positions within this continuous stream. This paradigm therefore clearly provides a more difficult task, in which non-adjacent syllable computations are apparently extremely difficult for learners to conduct.[4]

Does this result mean that the statistical learning mechanism we have discovered in previous work is limited to computing patterns only (or primarily) among immediately adjacent sound elements? There is another possibility: Learners might be able to compute certain types of non-adjacent regularities, but not others. We therefore considered the types of non-adjacent regularities that natural languages do (and do not) commonly exhibit.

In fact, the type of non-adjacent regularity we had built into the languages we were studying does not commonly occur in natural languages.[5] This regularity involves a patterned relationship between syllables 1 and 3 of a word, where 1 and

[4] Pena, Bonatti, Nespor, and Mehler (2002) report successful learning of non-adjacent dependencies using our paradigm. However, we believe their result was obtained through a phonetic cue to grouping that was irrelevant to the point of their own study, but would be a confound that we were careful to exclude in our own work. In particular, their 1–X–3 words contain stop consonants in the initial and final syllables of every word, but contain liquids and fricatives in the medial syllables of every word. When we include the same phonetic cue, we obtain the same results they did: subjects learn the statistical grouping. However, when we exclude this artifact or rearrange the syllables so that this phonetic contrast is no longer systematically correlated with word structure, our subjects no longer learn the non-adjacent dependency (and instead perform as reported in the present studies). To repeat, this point is not pertinent to the argument made by Pena et al., but in the present study would be a confound that we have carefully avoided. This difference in results is not due, we believe, to differences between English and French speakers (which could arise, for example, from differences in the prominence of syllables for segmentation in the two languages). We have run 7 native speakers of French on our non-adjacent syllables task of Experiment 2 and find the same failure to learn as exhibited by native English speakers.

[5] We are grateful to Katherine Demuth for helpful discussion of this issue.

3 never occur immediately adjacent to one another, and where the privileges of syllable position 2 are the same for all the different words in the language.[6] What particular aspect of this structure might be unusual or difficult to learn?

One possibility is that this difficult non-adjacent regularity occurs among *syllables*—relating elements of like kind, while skipping over other elements of the same kind. In contrast, in languages like Hebrew and Arabic, words are built out of patterns among non-adjacent phonemic *segments*, with word stems formed out of segments of one kind (consonants), while the intervening segments are of another kind (vowels). Perhaps either the difference in the type of element, or the difference in the contrast between patterned and intervening elements, might create a difference in learnability. To test this hypothesis, we built new miniature languages, structured more like those of Hebrew and Arabic, and contrasted the learning of such languages with the languages we had previously been studying.

4. Experiment 2: Non-adjacent syllables versus non-adjacent phonemic segments

In the present experiment, we built languages with two different types of non-adjacent regularities, but with other aspects of their structure fairly similar. One type of language involved *non-adjacent syllables*, like the languages we studied in the experiments described above, with transitional probabilities of 1.0 between the first and third syllables of a 3-syllable sequence, while the intervening second syllable varied. In contrast, the second type of language involved patterned regularities among *non-adjacent phonemic segments*. In this type of language, we created transitional probabilities of 1.0 among the consonants of a 3-syllable sequence, while the vowels that intervened between these consonants varied. The two types of languages were similar in the inventory of sounds used, the length of the words of the language (all 3-syllable), and the magnitudes of the high and low transitional probabilities that defined the word structure (1.0 among the patterned non-adjacent elements, .5 or lower among the adjacent and irrelevant non-adjacent elements). Then we asked whether this second type of language—common in natural languages of the world, such as Hebrew and Arabic—could be easily learned by our subjects (who were monolingual native speakers of English), whereas the first type of language—not common in natural languages of the world—would continue to be difficult to acquire.

[6] Natural languages do not commonly construct words out of a non-adjacent syllable pattern; most phonological patterns occur among adjacent phonetic segments or syllables. In some languages, such as Tagalog, words may contain 2-syllable stems, with certain inflections that can be inserted between the two syllables. But the two syllables also often occur adjacent to one another. Moreover, in languages that exhibit such patterns, the non-adjacent forms are relatively difficult for children to acquire and are generally acquired much later than comparable adjacent forms (Slobin, 1973).

4.1. Method

4.1.1. Subjects

Forty-seven undergraduates were recruited from Brain and Cognitive Sciences classes at the University of Rochester. Subjects were paid $6 for their participation. All subjects were monolingual English speakers, and none had previously been in a statistical learning experiment. All subjects were exposed to a single language, whose structure was defined by non-adjacent syllables or non-adjacent segments (consonants), and then tested using a 2AFC task comparing words versus partwords, as described below. Each type of language was produced and tested in two different phonetic instantiations, called Language A and Language B, in order to guard against idiosyncratic preferences for particular sounds, syllables, or their combinations. The final sample consisted of 12 participants in each of the two non-adjacent syllable languages, and 13 and 10 participants in the two non-adjacent segment languages.

4.1.2. Stimulus materials

All four languages consisted of the same inventory of six consonants (b, p, d, t, g, k) and six vowels (a, i, u, e, o, ae) which, when combined, rendered an inventory of 8 CV syllables in the non-adjacent syllable languages (di, ki, tae, gu, po, ga, ke,bu) and 12 CV syllables in the non-adjacent segment languages (pa, gi, tae, gu, te, po, da, ki, ku, bae, bu, do). As shown in Table 2, the two non-adjacent syllable languages each consisted of 6 trisyllabic words, constructed from three unique syllable-frames and two intervening syllables that could occur in each frame. As shown in Table 3, the two non-adjacent segment languages each consisted of 16 trisyllabic words, constructed from two unique consonant-frames and two vowels that could follow each consonant.

Table 2
Design of two non-adjacent syllable languages used in Experiment 2

CV_1 $[CV_2]$ CV_4
$[CV_3]$

CV_5 $[CV_2]$ CV_6
$[CV_3]$

CV_7 $[CV_2]$ CV_8
$[CV_3]$

1st–3rd Syllable word-frames	2nd Syllables
Language A	
di_tae	ki
po_ga	gu
ke_bu	
Language B	
bae_ku	pa
te_da	be
go_pi	

Table 3
Design of two non-adjacent segment (consonant) languages used in Experiment 2

$C_1[v_1]$	$C_2[v_3]$	$C_3[v_5]$
$[v_2]$	$[v_4]$	$[v_6]$
$C_4[v_1]$	$C_5[v_3]$	$C_6[v_5]$
$[v_2]$	$[v_4]$	$[v_6]$

Consonant-frames	Vowel-fillers
Language A	
p_g_t_	[_a] [_i] [_ae]
d_k_b_	[_o] [_u] [_e]
Language B	
t_d_k_	[_ae] [_a] [_i]
b_p_g_	[_e] [_o] [_u]

A continuous stream of words for each of the four languages was created by generating a randomized list using similar methods to those of Experiment 1. For the two non-adjacent syllable languages, 5 blocks, each consisting of a different random ordering of 2 tokens of each of the 6 words, were concatenated into a text in 15 different random orders, with the stipulation that the same word, and the same word-frame, never occurred twice in a row. All word boundaries were removed from the text, rendering a list of 2700 syllables. The text was then read by the MacInTalk speech synthesizer, using the text-to-speech application Speaker, running on a Power Macintosh G3 computer. Because the synthesizer was not informed of word boundaries, it did not produce any acoustic word boundary cues and produced equivalent levels of coarticulation between all syllables. The speech stream contained no pauses and was produced by a synthetic female voice (Victoria) in monotone. The output of the synthesizer was recorded to audiotape from the sound output of the Power Macintosh computer and then recorded again into SoundEdit 16 version 2. Once recorded in SoundEdit, each syllable was edited to .20–.22 s in length, in order to ensure that there were no differences in the length of syllables that could correlate with the position of syllables in the words. This stream of speech contained no pauses and played at a rate of 294 and 293 syllables per minute (for Language A and B, respectively). It was recorded from the sound output of the Power Macintosh onto audiocassette tape, for later presentation to the subjects. The tape consisted of a 9- to 9 1/2-min stream of speech which was repeated twice, for a total of 19 min of exposure.

The same procedure was used to generate the streams for the non-adjacent segment languages, except that these streams were formed from 6 blocks, each consisting of a constrained random ordering of one token of each of the 16 words in the language. These 6 blocks were concatenated into a text in 10 different random orders, rendering a list of 2880 syllables. All randomization was done with the stipulation that the same word never occurred twice in a row, and (in order to control adjacent syllable statistics at word boundaries) each word-final syllable could only be followed by either of two particular word-initial syllables. After recording and editing as described above, the stream of speech contained no pauses and played at a rate of 278 and 280 syllables per minute (for Language A and B, respectively). It was recorded from the sound output of the Power Macintosh onto audiocassette tape, for later presentation to the subjects. The tape consisted of a 10-min stream of speech which was repeated twice, for a total of 20 min of exposure.

For the non-adjacent syllable languages, the resultant statistical structure was similar to Experiment 1 (though simplified by having only 3 frames and 2 middle syllables), in which the only available information for extracting words was the greater statistical regularity of non-adjacent syllable sequences within words than of syllable sequences that spanned a word boundary. As in Experiment 1, this grouping could not be learned by computing transitional probabilities only among *adjacent* syllables: the adjacent transitional probabilities within words consisted of a transitional probability of .50 between the first and second syllables within a word, .33 between the second and third syllables within a word, and .50 between the last syllable of a word and the first syllable of the following word. In contrast, the *non-adjacent* statistical structure was a transitional probability of 1.0 between the first and third syllable within a word, compared with a transitional probability of .33 between the second syllable of one word and the first syllable of the next word and .50 between the third syllable of one word and the second syllable of the next word. Thus, the predictability of adjacent syllables throughout the entire speech stream was low (TP = .33–.50), and the predictability of non-adjacent syllables that spanned a word boundary was also low (TP = .33–.50), whereas the predictability of non-adjacent syllables *within* a word was high (TP = 1.0).

In comparison, for the non-adjacent segment languages, the only available information for extracting words was the greater statistical regularity of non-adjacent *segment* sequences within words. Each language was comprised of two 3-consonant frames, with 2 different vowels possible in each of the vocalic positions. Given this type of structure, the transitional probabilities between the consonants within a word were 1.0; the transitional probabilities between the consonants across word boundaries were .5. However, the word structure and the stream ordering rules were designed so that no adjacent transitional probability computation, either between adjacent syllables or between adjacent segments, would produce a coherent grouping. The transitional probabilities between *adjacent syllables* within words were .5. In order to make the transitional probabilities between adjacent syllables across word boundaries also equal to .5, the speech streams for these languages were created with a constrained randomization rule (each word-final syllable could only be followed by either of two word-initial syllables). The transitional probabilities between *adjacent segments* (from consonant to vowel and vowel to consonant) were also .5 all along the stream. In short, then, given a stream of words following these patterns, there is no grouping of syllables into words if adjacent syllable or adjacent segment relations are computed. However, words can readily be learned if non-adjacent (consonant) segment regularities are computed.

Learning the statistical coherence of the non-adjacent syllable or non-adjacent segment patterns within words was tested by asking whether subjects could discriminate words from partwords. Partwords were trisyllabic sequences that spanned a word boundary (and therefore had occurred in the exposure stream). Partwords were of two types: (a) a 3–1–2 pattern, consisting of the last syllable of one word and the first two syllables of another word, and (b) a 2–3–1 pattern, consisting of the last two syllables of one word and the first syllable of another word.

Table 4 lists the words and partwords for all 4 languages. In the two non-adjacent syllable languages, all six words were paired exhaustively with six partwords (three of the 3–1–2 type and three of the 2–3–1 type) to form 36 word/partword test items. Words and partwords were fairly similar in *adjacent syllable* statistics: the adjacent syllables in the words had transitional probabilities of .50–.33, whereas adjacent syllables in the partwords had transitional probabilities of .50–.50 or .33–.50 (for the 3–1–2 and 2–3–1 types, respectively). In contrast, the *non-adjacent syllable* statistics between the first and third syllables were high in the test words (TP = 1.0) and low in the partwords (TP = .50 or .33 for the 3–1–2 and 2–3–1 types, respectively).

In the non-adjacent segment languages, four of the 16 words were paired exhaustively with four partwords (two of the 3–1–2 type and two of the 2–3–1 type) to form 16 word/partword test pairs, which were each presented twice in different orders for a total of 32 test items. Words and partwords were identical in their *adjacent segment* and *adjacent syllable* statistics: words and partwords both had transitional probabilities of .50 between adjacent segments and adjacent syllables. In contrast,

Table 4
Test words and test partwords used in Experiment 2

	Words	Partwords
Non-adjacent syllable languages		
Language A	di ki tae	tae po ki
	di gu tae	ga ke ki
	po ki ga	bu di gu
	po gu ga	gu tae di
	ke ki bu	ki ga ke
	ke gu bu	gu bu po
Language B	bae pa ku	ku te be
	bae be ku	da go be
	te pa da	pi bae pa
	te be da	be ku go
	go pa pi	pa da bae
	go be pi	pa pi te
Non-adjacent segment languages		
Language A	do ki bae	be po gu
	da ku be	tae da ki
	po gu tae	ku bae pa
	pa gi te	gi te do
Language B	bae po gu	gi tae da
	be pa gi	ku be po
	tae da ku	pa gu te
	te do ki	do ki bae

the *non-adjacent segment* statistics (between the consonants) were high within words (TPs = 1.0 and 1.0) but low across the word boundary in the partwords (TP = .50).

Each of the test words and partwords was generated separately by the MacInTalk speech synthesizer, in the same way as described for the streams above, except that each was generated in isolation. This produced a falling intonation on the final syllable of each item, making each (words and partwords) sound like a word spoken in isolation. A test item consisted of a test word paired with a partword, in counterbalanced order, and was recorded to audiocassette tape with a 1 s silent interval between each word/partword that formed a test pair and a 5 s silent interval between each test pair. Two different randomized orders of the test items were constructed and presented, counterbalanced across subjects.

4.1.3. Apparatus

The experiment was conducted in an IAC sound-attenuated booth. The recorded speech stream was presented to subjects using an Onkyo stereo cassette tapedeck TA-2140 and a Proton 301 amplified speaker system. Test stimuli were presented using the same equipment, and subjects indicated their judgments on a pre-printed answer sheet.

4.1.4. Procedure

The procedure was the same as in Experiment 1, but with slightly different instructions. Subjects were instructed to listen to a "nonsense" language. They were not informed about the structure of the language; they were told simply to listen to the language, after which they would be tested on what they had learned. They were informed that the listening phase of the experiment consisted of two short segments, followed by a test. They were given a short break after the first 9–10-min listening segment. All subjects were run individually. After a total of 18–20 min of listening, subjects received a two-alternative forced-choice test. Two different randomized orders of presentation of the test items were used, counterbalanced across subjects.

4.2. Results

Fig. 3 presents the results on the 2AFC test for the 2 instantiations of the non-adjacent syllable languages (on the left) and for the 2 instantiations of the non-adjacent segment languages (on the right). In both cases the 2AFC test asked subjects to choose between words of the language (tri-syllabic sequences following the respective patterns of the language and capturing the non-adjacent transitional probabilities of 1.0 that had been displayed in the exposure stream), as compared with partwords (tri-syllabic sequences that had also occurred in the exposure stream but fell across the word boundaries defined by the transitional probability groupings).

As before, *non-adjacent syllable* regularities were not acquired. For the two non-adjacent syllable languages, performance on the 2AFC test did not exceed chance; in fact, in both cases performance was significantly *below* chance (Language A: mean = 13.42 out of 36, or 37.27% correct, $t(11) = -2.57$, $p = .03$; Language B: mean = 14.92 out of 36, or 41.44% correct, $t(11) = -2.48$, $p = .03$; overall:

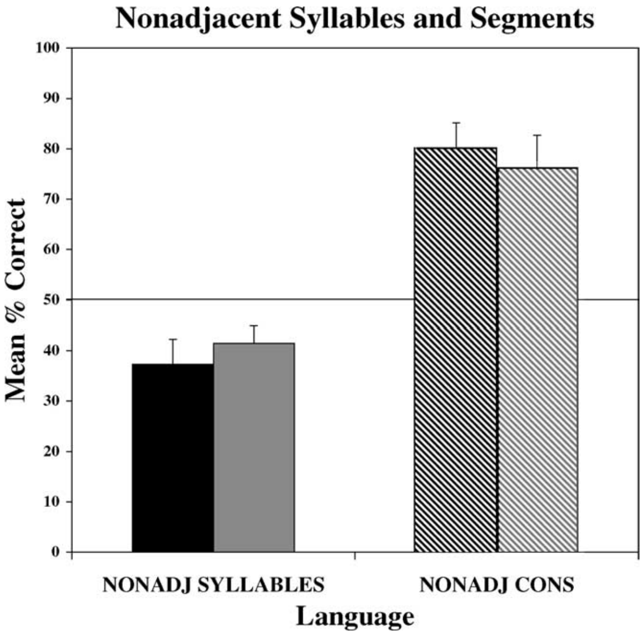

Fig. 3. Mean percent correct on the non-adjacent syllable languages (on the left) and on the non-adjacent segment (consonant) languages (on the right). For each language type, one bar shows the results on Language A, and the other shows the results on Language B.

mean = 14.17 out of 36, or 39.35% correct, $t(23) = -3.56$, $p = .002$). In contrast, *non-adjacent segment* regularities were readily acquired. For the two non-adjacent segment languages, performance on the 2AFC test significantly and substantially exceeded chance (Language A: mean = 25.62 out of 32, or 80.05% correct, $t(12) = 5.88$, $p < .0001$; Language B: mean = 24.40 out of 32, or 76.25% correct, $t(9) = 4.09$, $p = .0027$; overall: mean = 25.09 out of 32, or 78.40% correct, $t(22) = 7.20$, $p < .0001$). A two-way ANOVA testing Language Type × Instantiation showed a significant main effect of Language Type ($F(1, 43) = 59.72$, $p < .0001$) and no other significant effects.

4.3. Discussion

As we found in our previous studies, learners do not readily acquire regularities among non-adjacent syllables. However, learners do quite readily acquire regularities among non-adjacent segments (consonants). This contrast shows that human learners are not limited to learning adjacent regularities and are not incapable of computing non-adjacent regularities across the board. Rather, subjects in our experiments display a striking selectivity in the type of non-adjacent regularities they are able to learn.

Subjects did not learn the non-adjacent syllable languages. In contrast, they learned the non-adjacent segment languages quite easily, despite the fact that the consonant patterns these languages exhibited are not a widespread pattern in their native language (English).[7] These findings support the hypothesis that certain types of regularities—those that are common in natural languages—are readily learned, while others—those that are uncommon in natural languages—are not. In turn, these findings suggest that these patterns in natural languages may arise, at least in part, because of the selectivities of learning.

Before discussing these hypotheses further, however, we sought one more check on our results. Despite our many attempts to find evidence that non-adjacent syllable languages could be learned, there is one final possibility that could have prevented learning in our previous studies, which we investigated in a follow-up experiment. In Experiment 2 we attempted to match the two types of languages in many ways, or to favor non-adjacent syllable languages where complexity was not matched. However, in one regard our previous non-adjacent syllable languages might be viewed as more complicated to acquire than our non-adjacent segment languages. In particular, while there were more words in the non-adjacent segment languages (16) than in the non-adjacent syllable languages (6), one could argue that the relevant dimension of language complexity is the number of frames (segments or syllables) rather than the number of words. The non-adjacent segment languages in Experiment 2 had *two* word-frames (two 3-consonant frames, with two different vowels intervening), whereas our simplest non-adjacent syllable languages had *three* frames (three 2-syllable frames, with two different syllables intervening; see Tables 2–4). Experiment 2A was conducted to equate the number of frames between the two types of languages. In this experiment we used an even simpler version of non-adjacent syllable languages, in which there were only two non-adjacent syllable frames, with two intervening middle syllables, producing a total of four words in the language.

5. Experiment 2A: Control for the number of syllable frames

The structure of the non-adjacent syllable languages used in this experiment was identical to that used in Experiment 2, except that the number of word-frames was reduced from three to two. This also resulted in a reduction in the number of words in the language, from six to four. By the metric of syllable frames, then, these languages were equal in simplicity to the non-adjacent segment languages. By other metrics—for example, the number of total words in the language—these languages are much simpler than the non-adjacent segment languages previously studied.

[7] English displays some scattered examples of both the non-adjacent syllable and the non-adjacent segment patterns, but uses neither as a regular part of word formation processes. None of the subjects in our experiments had been exposed to languages that did display these patterns; all were monolingual native speakers of English.

5.1. Method

5.1.1. Subjects

Sixteen undergraduates were recruited from Brain and Cognitive Sciences classes at the University of Rochester and were paid $7.50 for their participation. All subjects were monolingual English speakers, and none had previously been in a statistical learning experiment. Each subject was exposed to one non-adjacent syllable language. This structure was produced in two different phonetic instantiations, Language A and Language B, to guard against idiosyncratic preferences for particular sounds, syllables, or their combinations.

5.1.2. Stimulus materials

Both languages consisted of the same inventory of six consonants (b, p, d, t, g, k) and five of the six vowels (a, i, u, e, o) used in Experiment 2. These segments were combined to form six unique syllables in each of the two languages (A: do, ke, ki, gu, ta, bu; B: te, go, pa, be, da, pi). As shown in Table 5, the two non-adjacent syllable languages each consisted of 4 trisyllabic words, constructed from two unique syllable frames and two intervening syllables that could occur in each frame.

The stimulus streams were generated and edited in the same way as in Experiment 2. Words were combined into blocks comprised of various random orderings of words, which were then concatenated into a continuous text, with the stipulation that the same word never occurred twice in a row. As in the non-adjacent segment languages of Experiment 2, in order to control adjacent syllable statistics at word boundaries, each word could only be followed by either of two other words (words 1 and 4 could only be followed by words 2 or 3, and vice versa). This resulted in a 'flat' set of adjacent syllable transitional probabilities (.50) across the entire stream. As in Experiment 2, the text was read by the MacInTalk speech synthesizer, recorded into SoundEdit 16 version 2, and then edited to produce equal duration syllables and words and no acoustic cues to word boundaries.

Table 5
Design and test words and partwords used in the non-adjacent syllable languages of Experiment 2A

CV_1 [CV_2] CV_4
[CV_3]

CV_5 [CV_2] CV_6
[CV_3]

	Words	Partwords
Language A	do ki ta	ki bu do
	do gu ta	gu bu do
	ke ki bu	ta ke ki
	ke gu bu	ta ke gu
Language B	te pa da	pa pi te
	te be da	be pi te
	go pa pi	da go pa
	go be pi	da go be

The resultant statistical structure of the two languages was very similar to Experiment 2, with the transitional probability between non-adjacent syllables within words equal to 1.0, and all other transitional probabilities (between adjacent or non-adjacent syllables) equal to .50. Learning the statistical coherence of the non-adjacent syllable structure within words of the languages was assessed by asking whether subjects could discriminate words from partwords. All four words were paired exhaustively with four partwords (two of the 3–1–2 type and two of the 2–3–1 type) to form 16 word/partword test pairs, which were each presented twice, in different orders, for a total of 32 test items. Words and partwords were identical in their *adjacent* syllable statistics (both had adjacent transitional probabilities of .50–.50), whereas the *non-adjacent* syllable statistics of the first and third syllables of the words (1.0) were higher than those of the partwords (.50).

5.1.3. Apparatus and procedure

These were the same as in Experiment 2.

5.2. Results and discussion

Fig. 4 presents the results of the 2AFC test for the two 2-frame non-adjacent syllable languages in Experiment 2A, as well as the results of the non-adjacent syllable

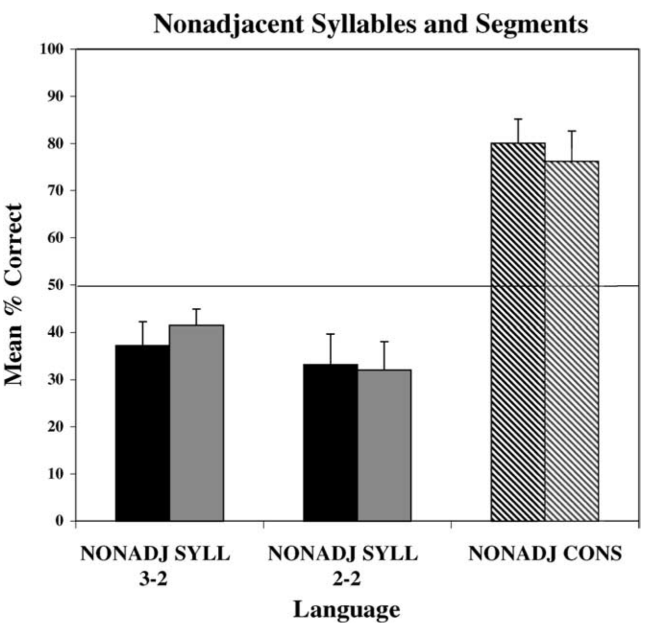

Fig. 4. Mean percent correct on the non-adjacent syllable languages in Experiment 2 (on the left) and Experiment 2A (in the middle), as compared with the non-adjacent segment (consonant) languages in Experiment 2 (on the right). For each language type, one bar shows the results on Language A and the other shows the results on Language B.

and segment languages from Experiment 2. As in the non-adjacent syllable languages of Experiment 2, the performance on the non-adjacent syllable languages in the present experiment did not exceed chance; in fact, for both Language A and Language B, performance was significantly *below* chance (Language A: mean = 10.62 out of 32, or 33.2% correct, $t(7) = -2.60$, $p < .04$; Language B: mean = 10.25 out of 32, or 32.0% correct, $t(7) = -3.00$, $p < .02$; overall: mean = 10.44 out of 32, or 32.6% correct, $t(15) = -4.08$, $p = .001$). Although it is not clear why subjects in both Experiments 2 and 2A showed below-chance performance on the 2AFC tests in the non-adjacent syllable conditions, they clearly did not learn the coherent statistical structure of the 3-syllable words. Moreover, their performance contrasts strikingly with the ease of learning shown by subjects in the non-adjacent *segment* languages.

In short, then, non-adjacent syllable regularities are not readily learned in our paradigm, even when they are extremely simple. In contrast, non-adjacent segment regularities, at least among consonants, are quite readily learned. Before concluding that there is a fundamental contrast between nonadjacent patterns of two types, however, we turn to one more case of interest. In Experiment 2, we investigated only one of the types of non-adjacent segment regularities that occurs commonly in natural languages: regularities among consonants, skipping over vowels. In our final experiment we examine another: regularities among vowels, skipping over consonants. This type of pattern occurs, for example, in Turkish, where words exhibit what is called 'vowel harmony': within a word, certain features of the vowels across syllables of the word are made to match or 'harmonize' with one another. This type of patterning means that young learners can potentially use the regularity of vowels to determine which sound sequences form words. In our experiment, however, we introduced a simpler pattern among vowels—in fact, the precise complement of the one built among consonants in Experiment 2. The experiment asks whether this type of non-adjacent segment regularity is easily learned, like the one we studied in Experiment 2, and therefore whether the larger generalization about learnability indeed contrasts non-adjacent segments (both consonants and vowels) with non-adjacent syllables.

6. Experiment 3: Non-adjacent phonemic segments (vowels)

In this experiment we built a new type of language with patterned regularities among non-adjacent phonemic segments: this time among the vowels, skipping over the consonants. In this type of language, we created transitional probabilities of 1.0 among the vowels of a 3-syllable sequence, while the consonants that intervened between these vowels varied. These languages were similar to the non-adjacent syllable and non-adjacent segment languages of Experiment 2 in other ways: in the inventory of sounds used, the length of the words of the language (all 3-syllable), and the magnitudes of the high and low transitional probabilities that defined the word structure (1.0 among the patterned non-adjacent elements, .5 or lower among the adjacent and irrelevant non-adjacent elements). The present experiment thus extends the question of Experiment 2, asking whether another type of non-adjacent segment regularity, also common in languages of the world, will be easy to acquire.

6.1. Method

6.1.1. Subjects

Twenty-six undergraduates were recruited from Brain and Cognitive Sciences classes at the University of Rochester. Subjects were paid $6 for their participation. All subjects were monolingual English speakers, and none had previously been in a statistical learning experiment. All subjects were exposed to a single language, whose structure was defined by non-adjacent segments (vowels), and then tested using a 2AFC task comparing words versus partwords, as described below. Each type of language was produced and tested in two different phonetic instantiations, called Language A and Language B, in order to guard against idiosyncratic preferences for particular sounds, syllables, or their combinations. The final sample consisted of 12 and 14 participants in each of the two non-adjacent segment languages.

6.1.2. Stimulus materials

Both languages consisted of the same inventory of six consonants (b, p, d, t, g, k) and six vowels (a, i, u, e, o, ae) used in the non-adjacent segment (consonant) languages in Experiment 2. The languages were constructed from the same 12 CV syllables, but structured so that the vowels formed a consistent word-frame, while the consonants varied (see Table 6). The two non-adjacent (vowel) segment languages each consisted of 16 trisyllabic words, constructed from two unique 3-vowel frames and two different consonants possible in each of the consonantal positions.

A continuous stream of words for each of the languages was created by generating a randomized list, using the same methods as in the non-adjacent segment languages of Experiment 2. Six blocks, each consisting of a constrained random ordering of one token of each of the 16 words in the language, were concatenated into a text in 10 different random orders, rendering a list of 2880 syllables. All randomization was done with the stipulation that the same word never occurred twice in a row, and

Table 6
Design of two non-adjacent segment (vowel) languages used in Experiment 3

$[c_1]V_1$	$[c_3]V_2$	$[c_5]V_3$
$[c_2]$	$[c_4]$	$[c_6]$
$[c_1]V_4$	$[c_3]V_5$	$[c_5]V_6$
$[c_2]$	$[c_4]$	$[c_6]$

Vowel-frames	Consonant-fillers
Language A	
_a_u_e	[p_] [g_] [t_]
_o_i_ae	[d_] [k_] [b_]
Language B	
_ae_a_u	[t_] [d_] [k_]
_e_o_i_	[b_] [p_] [g_]

(in order to control adjacent syllable statistics at word boundaries) each word-final syllable could only be followed by either of two particular word-initial syllables. All word boundaries were removed from the text, which was then read by the MacInTalk speech synthesizer, using the text-to-speech application Speaker, running on a Power Macintosh G3 computer. Because the synthesizer was not informed of word boundaries, it did not produce any acoustic word boundary cues and produced equivalent levels of coarticulation between all syllables. The speech stream contained no pauses and was produced by a synthetic female voice (Victoria) in monotone. The output of the synthesizer was recorded to audiotape from the sound output of the Power Macintosh computer and then recorded again into SoundEdit 16 version 2. Once recorded in SoundEdit, each syllable was edited to .20–.22 s in length, in order to ensure that there were no differences in the length of syllables that could correlate with the position of syllables in the words. This stream of speech contained no pauses and played at a rate of 284 syllables per minute (for both Languages A and B). It was recorded from the sound output of the Power Macintosh onto audiocassette tape, for later presentation to the subjects. The tape consisted of a 10-min stream of speech which was repeated twice, for a total of 20 min of exposure.

As in the non-adjacent segment languages of Experiment 2, the only available information for extracting words from this stream was the greater statistical regularity of non-adjacent *segment* sequences within words. In contrast to Experiment 2, however, the non-adjacent segment regularities in the present languages occurred among the vowels. Each language was comprised of two 3-vowel frames, with 2 different consonants possible in each of the consonantal positions. Given this type of structure, the transitional probabilities between the vowels within a word were 1.0; the transitional probabilities between the vowels across word boundaries were .5. However, the word structure and the stream ordering rules were designed so that no adjacent transitional probability computation, either between adjacent syllables or between adjacent segments, would produce a coherent grouping. The transitional probabilities between *adjacent syllables* within words were .5. In order to make the transitional probabilities between adjacent syllables across word boundaries also equal to .5, the speech streams for these languages were created with a constrained randomization rule (each word-final syllable could only be followed by either of two word-initial syllables). The transitional probabilities between *adjacent segments* (from consonant to vowel and vowel to consonant) were also .5 all along the stream. In short, then, given a stream of words following these patterns, there is no grouping of syllables into words if adjacent syllable or adjacent segment relations are computed. However, words can readily be learned if non-adjacent (vowel) segment regularities are computed.

Learning the statistical coherence of the non-adjacent (vowel) segment patterns within words was tested by asking whether subjects could discriminate words from partwords. Partwords were trisyllabic sequences that spanned a word boundary (and therefore had occurred in the exposure stream). Partwords were of two types: (a) a 3–1–2 pattern, consisting of the last syllable of one word and the first two syllables of another word, and (b) a 2–3–1 pattern, consisting of the last two syllables of one word and the first syllable of another word.

Table 7 lists the words and partwords for both languages. Four of the 16 words were paired exhaustively with four partwords (two of the 3–1–2 type and two of the 2–3–1 type) to form 16 word/partword test pairs, which were each presented twice in different orders for a total of 32 test items. Words and partwords were identical in their *adjacent segment* and *adjacent syllable* statistics: words and partwords both had transitional probabilities of .50 between adjacent segments and adjacent syllables. The non-adjacent segment transitional probabilities between the consonants were also .50 throughout the speech stream. In contrast, the *non-adjacent segment* statistics between the vowels were high within words (TPs = 1.0 and 1.0) but low across the word boundary in the partwords (TP = .50).

Each of the test words and partwords was generated separately by the MacInTalk speech synthesizer, in the same way as described for the streams above, except that each was generated in isolation. This produced a falling intonation on the final syllable of each item, making each (words and partwords) sound like a word spoken in isolation. A test item consisted of a test word paired with a partword, in counterbalanced order, and was recorded to audiocassette tape with a 1 s silent interval between each word/partword that formed a test pair and a 5 s silent interval between each test pair. Two different randomized orders of the test items were constructed and presented, counterbalanced across subjects.

6.1.3. Apparatus and procedure

The apparatus and procedure were the same as in Experiment 2.

6.2. Results

Fig. 5 presents the results on the 2AFC test for the 2 instantiations of the new non-adjacent segments (vowels) languages (on the right), along with the data from the other two language types of Experiment 2: the non-adjacent syllables languages (two pairs of bars on the left) and the non-adjacent segments (consonants) languages (in the middle). In all cases the 2AFC test asked subjects to choose between words of the language (tri-syllabic sequences following the respective patterns of the language

Table 7
Test words and test partwords used in Experiment 3

	Words	Partwords
Non-adjacent segment languages		
Language A	pa ku te	te do ki
	da gu be	bae pa gu
	po gi tae	ku be po
	do ki bae	gi tae da
Language B	tae pa ku	ku be po
	bae da gu	gi tae da
	te do ki	pa gu te
	be po gi	do ki bae

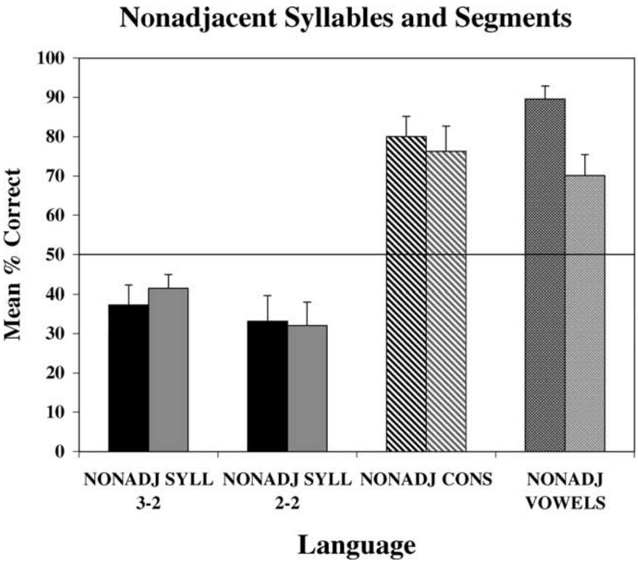

Fig. 5. Mean percent correct on the non-adjacent syllable languages (two pairs of bars on the left) and on the non-adjacent segment (consonant and vowel) languages (two pairs of bars on the right). For each language type, one bar shows the results on Language A, and the other shows the results on Language B.

and capturing the non-adjacent transitional probabilities of 1.0 that had been displayed in the exposure stream), as compared with partwords (tri-syllabic sequences that had also occurred in the exposure stream but fell across the word boundaries defined by the transitional probability groupings).

The new non-adjacent segment regularities among vowels were readily acquired. For both of the two language instantiations, performance on the 2AFC test significantly and substantially exceeded chance (Language A: mean = 28.67 out of 32, or 89.58% correct, $t(11) = 11.78$, $p < .0001$; Language B: mean = 22.43 out of 32, or 70.09% correct, $t(13) = 3.75$, $p = .0024$; overall: mean = 25.31 out of 32, or 79.09% correct, $t(25) = 7.74$, $p < .0001$). These results are therefore similar to those for the non-adjacent segment regularities among consonants, and quite different from those for the non-adjacent syllable regularities.

6.3. Discussion

The present experiment has shown that subjects are readily able to acquire regularities among non-adjacent vowels, just as they were able to acquire regularities among non-adjacent consonants. In contrast, as shown in Experiments 1 and 2, regularities among non-adjacent syllables are not readily acquired. These differences hold up even though, in other ways, the contrasting miniature languages are very similar.

In sum, then, the findings support the hypothesis that certain types of regularities—those that are common in natural languages—are readily learned, while others—those that are uncommon in natural languages—are not. We turn next to a discussion of the findings across our experiments, candidate explanations for why these selectivities of learning might occur, and the implications of these findings for natural language structure and acquisition.

7. General discussion

The aim of the present experiments was to investigate learners' ability to acquire non-adjacent regularities among speech sounds. In previous work we have demonstrated that human learners have a remarkable capacity to compute complex co-occurrence statistics among speech sounds (as well as other types of auditory stimuli), and to do so rapidly, online, and simultaneously over a fairly large number of sounds across a continuous stream of speech (Aslin et al., 1998; Newport & Aslin, 2000; Saffran et al., 1996a, 1996b; Saffran, Johnson, Aslin, & Newport, 1999). However, our previous studies left two important questions unanswered. First, in all of these studies, the patterns on which learners were tested involved statistical regularities among immediately adjacent speech sounds. In contrast, natural languages contain not only immediately adjacent regularities, but also regularities among elements not immediately adjacent to one another (Chomsky, 1957). In order to understand how this statistical learning capacity might be employed in natural language learning, it is therefore necessary to ask whether it is limited to computing adjacent regularities, or rather whether it includes the ability also to compute non-adjacent regularities. If statistical learning were limited to adjacent regularities, such a mechanism would have to combine with another type of device, able to handle non-adjacent patterns, to successfully learn languages. On the other hand, if statistical learning operates on non-adjacent as well as adjacent elements, a further question concerns which types of computations on non-adjacent elements can be performed. While a wide range of adjacent regularities appears throughout natural languages, the types of non-adjacent regularities that languages exhibit are, interestingly, quite constrained; indeed, a major focus of modern linguistics has been to state in a principled way what these constraints are. A tantalizing question, then, is whether some of the constraints on non-adjacent regularities in natural languages might match with, and indeed arise from, constraints on the non-adjacent computations that human learners are able to perform.

A second, interrelated question concerns the units or elements on which statistical learning is performed. In our previous studies we have described the statistical regularities that form 'words' in the speech stream as comprised of high transitional probabilities among neighboring syllables (Saffran et al., 1996a, 1996b). However, our materials also included high transitional probabilities among neighboring phonemic segments; in these early studies we purposely built miniature languages in which both syllable and segment regularities would predict word forms. While these studies thus demonstrate that listeners can keep track of statistics in neighboring speech sounds, they do not distinguish whether the computations are performed on syllables, segments, or both.

In the present studies we have begun to address both of these additional questions. In these experiments we have constructed materials in which learners can extract word-like groupings only by computing various types of non-adjacent regularities; adjacent elements are controlled so that no grouping occurs if only adjacent regularities are computed. Across experiments and experimental conditions, different types of non-adjacent regularities are compared. Our results across these studies clearly show that adult learners are highly selective in the types of non-adjacent regularities they are readily able to compute. Non-adjacent *syllable* regularities are very difficult to acquire; while they can be performed in other studies employing simpler learning paradigms, they are not acquired in our studies. In contrast, non-adjacent *segment* regularities—patterns among consonants, skipping over vowels, and also patterns among vowels, skipping over consonants—are extremely easy to acquire and are readily learned in our paradigm. This contrast holds true even when the miniature languages are quite similar in the sounds employed, the magnitude and type of statistics required for learning, and the like.

Why might there be this difference in the ease of learning different types of non-adjacent regularities? There are several possibilities—some that can be eliminated on the basis of the present results, but others that will require future experimentation.

One class of possibilities we believe we can dismiss is that these differences in learning might arise from differences in the overall complexity of the languages or the differential availability of strategies based on computations among adjacent elements. With regard to complexity: while it is difficult to assess the overall complexity of patterns that differ in many ways, our non-adjacent syllable languages and our non-adjacent segment languages seem similar, or if anything, show greater complexity in the non-adjacent segment languages (which are more easily learned). Both types of languages have transitional probabilities of 1.0 between the relevant non-adjacent elements inside words, and transitional probabilities of .33–.5 between the non-adjacent elements across words and the adjacent elements within and across words. The non-adjacent syllable languages have 4 or 6 different words, formed out of 2 or 3 syllable-frames, while the non-adjacent segment languages have 16 different words, formed out of 2 segment-frames. Both types of languages were constructed from the same inventory of sounds and were synthesized and edited in approximately the same ways. Moreover, the use of two language instantiations per language type (and the consistency of our findings across the two language instantiations in each experiment) insures that our results are due to the type of structure these languages exhibit, and not to the details of which sounds are assigned to particular positions within the word.

While we have described our languages in terms of transitional probabilities among adjacent and non-adjacent elements, we should note that, in all of our languages, words also differ from partwords in the frequency with which these tri-syllabic sequences occur in the speech stream.[8] Is it possible, then, that the learning

[8] Aslin et al. (1998) demonstrated that infants can discriminate words from partwords even when these tri-syllabic frequencies are matched. But because of certain undesirable aspects of the experimental design required to do such matching—individual syllable frequencies and tri-syllabic frequencies of untested items are not matched—we do not use such a design in most of our studies.

in these studies occurs by keeping track of the frequency of bi- or tri-syllabic sequences, perhaps avoiding non-adjacent computations altogether? Or is it possible that the differential learning of different types of languages in these studies is due to the differential availability of such frequencies? The answer to this question is very likely no. All the languages we have examined here, both non-adjacent syllable and non-adjacent segment types, exhibit a difference in tri-syllabic sequence frequencies for words versus partwords. If this were the statistic used by subjects during learning, then, both types of languages should have been learned. Indeed, by this metric the non-adjacent syllable languages should be easier to learn than the non-adjacent segment languages, since the non-adjacent syllable languages are made up of only 4 or 6 trisyllabic sequences (in Experiments 2A and 2, respectively), whereas the non-adjacent segment languages are each made up of 16 trisyllabic sequences. The pattern of learning we have found across our language types thus suggests that it is not trisyllabic frequency, but rather the statistic on which our languages differ—transitional probabilities among various types of non-adjacent elements[9]—that learners are actually performing in the task.

Having eliminated, we believe, the more mundane explanations of our findings, we turn now to several more theoretically interesting accounts: first, the effects of element similarity on the ease/difficulty of computing non-adjacent regularities; second, differences in computing syllable versus segment regularities; and third, the interaction of distance and elements in statistical learning.

7.1. Element similarity and non-adjacent learning: Gestalt principles in statistical learning

One possible account of the relative ease of learning non-adjacent segment regularities, as compared to non-adjacent syllable regularities, is related to the effects of element similarity on the ease of learning. Our previous studies suggest that learning relationships between adjacent elements is relatively simple: we have demonstrated that human adults can learn relationships between a variety of types of adjacent elements, including speech sounds, tones, visual shapes displayed in temporal sequences, and serial motor responses (Creel, Newport, & Aslin, submitted; Fiser & Aslin, 2002; Hunt & Aslin, 2001; Saffran et al., 1996b, 1999), and also that human infants and tamarin monkeys can learn some of these same regularities (Hauser, Newport, & Aslin, 2001; Saffran et al., 1996a). In contrast, learning relationships between non-adjacent elements is relatively difficult. Experiment 1 and our various subsequent exploratory studies show that obtaining successful learning of non-adjacent regularities is not trivial. Comparable results appear in other paradigms. For example, the results of Gomez (2002), using a simple miniature syntax with only 3-word strings, indicate that adult and infant learners acquire relationships

[9] As noted earlier (see footnote 1), while we have frequently used the term *transitional probability* to describe our materials, we mean more technically to refer to any of a class of conditionalized statistics, including *mutual information* or *conditional entropy*. The present point is that our results suggest the relevant statistic could not instead be a frequency statistic (such as tri-syllabic co-occurrence frequency).

between the first and third words of these strings under only very specific circumstances. Marcus et al. (1999), also using 3-word strings, obtained learning of both an ABA and ABB regularity in infants; but a replication by Johnson (2002), using temporally ordered visual stimuli, finds the non-adjacent repetition (ABA) to be more difficult to learn. In the sequence learning literature, Cleeremans and McClelland (1991) and Cleeremans (1993) have shown that non-adjacent contingencies spanning identical embedded sequences (of 3 elements or more) are not learned by human learners and provide an especially difficult learning problem even for large SRNs. In our own recent work with musical tone streams (Creel et al., submitted), when two sets of melodies are temporally interleaved so that the patterned melodic regularities occur among non-adjacent tones, adult learners do not acquire these patterns, though they do learn the much less regular relationships between adjacent tones.[10]

Why, then, are non-adjacent *segment* regularities learned so easily? In both types of non-adjacent segment languages we have studied (Experiments 2 and 3), the non-adjacent segments that are regularly related to one another are of one element type, while the intervening segments are of another element type. In Experiment 2, the patterned segments were consonants (and the intervening unrelated elements were vowels); in Experiment 3, the patterned segments were vowels (and the intervening unrelated elements were consonants). An intuitive way of thinking about the ease of learning such regularities is that, when patterned elements are of one type, and the unrelated intervening elements are of another type, noticing and storing the regularities among elements of like kind is much easier, and the usual difficulty of non-adjacency is ameliorated.

Two theoretical frameworks provide possible explanations of this effect. One of these frameworks is the Gestalt principles of perception. According to the Gestalt principle of similarity (Wertheimer, 1938, 1944), we tend to perceive elements that are physically similar to one another as grouped together and more closely related than their objective temporal or spatial distances would suggest (see also Bregman, 1990, on auditory streaming). In our recent work examining musical tone streams (Creel et al., submitted), we have shown that these Gestalt principles of similarity and auditory streaming also constrain statistical learning. As noted above, when two sets of melodies are temporally interleaved, the regularities among non-adjacent tones are not readily acquired. But when the two sets of melodies are played in two different octaves, high versus low, and then temporally interleaved, the patterns among the temporally non-adjacent tones are readily acquired. The present results, with temporally interleaved consonants and vowels, are analogous to these findings, suggesting that Gestalt principles of similarity may also constrain statistical learning of speech.

A second framework, describing similar phenomena in more specifically linguistic terms, involves the notion of phonological *tiers* (Goldsmith, 1976, 1990; McCarthy, 1981). In the linguistic theory known as *autosegmental phonology*, phonological var-

[10] See footnote 2 for a discussion of the recent results of Pena et al. (2002).

iation in related word forms is captured by representing consonants and vowels on different tiers. By this contrast in representation, certain types of phonological processes (e.g., voicing assimilation, consonantal harmony) can apply to consonants, while others (e.g., vowel harmony) can apply to vowels. Consonants and vowels are placed in their interleaved order in syllables and words by associating the elements within tiers to a timing tier. Within this type of framework, the present results can be explained by suggesting that adjacency is not defined on surface phonetic forms, but rather is defined on more abstract representations, such as tiers. Consonants are adjacent to one another on one tier, while vowels are adjacent to one another on another tier. If statistical computations occur on this type of more abstract representation, then our non-adjacent segment languages may be easy to learn because they are in fact comprised of regularities among *adjacent* elements.

It is interesting to note that the two explanations we have offered—Gestalt and autosegmental—are quite similar, despite the very different frameworks from which they are derived.

7.2. Differences in computing syllable versus segment regularities

A different account of the ease of learning non-adjacent segment regularities, as compared to non-adjacent syllable regularities, concerns the type of element on which adult learners perform statistical computations during learning. As noted above, our earlier studies of statistical learning involved words that were formed from high transitional probabilities among both neighboring syllables *and* neighboring segments. These results therefore did not address the question of whether our subjects were, in fact, performing their computations on the syllables or on the segments of our speech streams. A possible interpretation of the present results is that adult learners perform their computations exclusively on segments and are unable to acquire statistical groupings based on syllables. On this interpretation, learners are able to compute non-adjacent regularities, but only among certain types of elements.

It is not likely that this limitation could be a basic perceptual one: a number of studies in the psycholinguistics literature suggest that human listeners can perceive both segments and syllables, depending on the task (Nygaard & Pisoni, 1995).[11] For example, adults can perform both phoneme and syllable monitoring tasks, and can learn to read both syllabary and alphabetic scripts. If there is a limitation to the statistical learning of segments, then, it would have to be a limitation on the locus of computations during learning. It is possible that human listeners perform statistical computations only on segments, or that they perform their computations initially on segments and then construct other statistics (e.g., the statistics of syllables and words) indirectly from these.

[11] Some investigators (Gleitman & Rozin, 1977; Mehler, 1981; Savin & Bever, 1970) have suggested that syllables are more accessible as perceptual units than are segments, though this claim is controversial. See Newport, Hauser, Spaepen, and Aslin (submitted), for further discussion.

A test of this hypothesis involves constructing two new types of languages: one in which the words are formed from adjacent syllable regularities (but cannot be learned through adjacent segment regularities), and one in which the words are formed from adjacent segment regularities (but cannot be learned through adjacent syllable regularities). This contrast is somewhat difficult to design, since typically, in natural languages as well as in our miniature languages, regularities of segments and of the syllables built out of these segments are correlated. But we have recently designed languages of these types and are in the process of determining which of them is readily learned.

7.3. The interaction of distance and elements: Hierarchical structure and statistical learning

A final possibility is that accounts one and two above—the nature of non-adjacency and distance, and the types of elements that human listeners perceive and compute—are both pertinent to understanding statistical learning. Extensive linguistic and psycholinguistic evidence suggests that speech is represented by human listeners in complex and linguistically structured ways. At the lowest level, speech may be represented in terms of features or acoustic transitions; segments, syllables, words, and phrases are formed from a series of hierarchically organized combinations of these smaller units. It is not yet clear where, at these various levels of representation, learners are capable of performing statistical computations, and among which of these types of units patterned regularities can be acquired. It is likely in such a perceptual system, however, that the organization of statistics across levels will be complex, and will favor a number of constraints on the types of computations that can be performed. One possibility, already noted, is that statistical computations may begin in terms of the smallest units of sound: keeping track of feature and/or segment combinations, for example. Evidence from our own studies, as well as from other research (Coady & Aslin, 2003; Maye, 2000; Maye, Werker, & Gerken, 2002; Vitevitch & Luce, 1998), suggests that both adults and children are sensitive to such statistics. Statistical regularities concerning syllable co-occurrence patterns may be computed separately, or may be assembled from the regularities of segments. In such a scheme, keeping track of the statistical regularities among syllables that are non-adjacent might be extremely indirect, involving greater (hierarchical) distance than keeping track of the statistical regularities among non-adjacent segments.

Much more research will be required to determine how these various types of computations are performed and assembled. A related set of interesting questions concerns how these phenomena are handled by infants, who have been hypothesized by some investigators (Bertoncini & Mehler, 1981; Jusczyk & Derrah, 1987; cf. Jusczyk, 1997, for discussion) to begin development with simpler and more holistic representations of speech and to develop the adult's more finely articulated representation of speech only with further maturation or learning. Addressed in the companion paper to the present article is how these phenomena are handled by non-human primates, who display some surprising abilities to process basic aspects of human speech streams (Hauser et al., 2001) but who would not be expected to exhibit the full array of human processing capabilities (see Newport, Hauser, Spaepen & Aslin, submitted). However our future studies turn out, though, the present results suggest that statistical learning by human adults is not limited to low-level computations on unstructured acoustic units, and raise the possibility that the nature of statistical learning interacts strongly with the types of elements and structured representations adults have as they process speech.

8. Conclusions

We believe that the present results provide a new and important step in understanding the nature of statistical learning and the ways in which it might be pertinent to the acquisition and structure of natural languages. The present studies asked whether statistical learning is limited to computations on adjacent sound sequences only, or rather whether learners can also perform computations on non-adjacent sound sequences. If a statistical learning mechanism could conduct its computations on adjacent sound sequences only, this might provide useful early data for learning but would have to be combined with another type of learning mechanism in order to analyze the more complex regularities characteristic of natural languages. On the other hand, if statistical learning can be performed on non-adjacent sound sequences, our previous results were only the first glimpse of a more elaborate mechanism. The present results suggest that statistical learning is not limited to computations on adjacent sound sequences. A question for further study is how complex and extensive statistical learning mechanisms are, and what types of computations comprise the set of statistical capabilities learners bring to the task of learning natural languages.

An important possibility suggested by the present results is that human statistical learning abilities are not limited to just a few elementary computations, but nonetheless are selective in ways that match the constraints that natural languages exhibit. In the present data, those types of non-adjacent regularities that human languages commonly exhibit are the same types that our learners readily acquire, while those types of non-adjacent regularities that human languages do not exhibit are also those that our learners do not readily acquire. This compatibility between learning and languages in turn suggests that natural language structures may be formed, at least in part, by the constraints and selectivities of what human learners find easy to acquire (Bever, 1970; Newport, 1981, 1982, 1990; Newport & Aslin, 2000). Our ongoing work is aimed at examining additional types of computations needed to acquire the further types of patterns human languages exhibit. The present results suggest that selectivities of learning might help to explain the universal aspects of these patterns, and therefore that elucidating the nature of statistical learning may play an important role in understanding natural languages.

Acknowledgments

We are grateful to Toby Calandra, Elizabeth Johnson, Kirti Sharma, Kelly Kinde, and Joanne Esse for their extensive help in conducting these experiments, to Marc Hauser, Jessica Maye, and an anonymous reviewer for comments on an earlier draft of the paper, to Alex Pouget for helpful discussion of entropy, and to Katherine Demuth for insightful discussion of nonadjacent dependencies in natural languages. This research was supported in part by NIH Grant DC00167 to ELN, NIH Grant HD37082 to RNA, and NSF Grant SBR-9873477 to RNA and ELN.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*, 321–324.

Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant perception. *Infant Behavior and Development, 4*, 247–260.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge: MIT Press.

Chomsky, N. A. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. A. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.

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

Chomsky, N. A. (1995). *The minimalist program*. Cambridge: MIT Press.

Cleeremans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing*. Cambridge: MIT Press.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120*, 235–253.

Coady, J. A., & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language, 30*, 441–469.

Creel, S. C., Newport, E. L., & Aslin, R. N. (submitted). Distant melodies: Statistical learning of non-adjacent dependencies in tone sequences (manuscript submitted).

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 458–467.

Gleitman, L. R., & Newport, E. L. (1995). The invention of language by children: Environmental and biological influences on the acquisition of language. In L. R. Gleitman & M. Liberman (Eds.), *Language* (Vol. 1). *An invitation to Cognitive Science* (2nd ed.). Cambridge: MIT Press.

Gleitman, L. R., & Rozin, P. (1977). The structure and acquisition of reading I: Relations between orthographies and the structure of language. In A. Reber & D. Scarborough (Eds.), *Toward a psychology of reading*. Hillsdale, NJ: Erlbaum.

Goldsmith, J. (1976). Autosegmental phonology. Unpublished doctoral dissertation, MIT.

Goldsmith, J. (1990). *Autosegmental and metrical phonology*. Oxford: Basil Blackwell.

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

Harris, Z. S. (1955). From phoneme to morpheme. *Language, 31*, 190–222.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition, 78*, B53–B64.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: Simultaneous extraction of multiple statistics. *Journal of Experimental Psychology: General, 130*, 658–680.

Johnson, S. P. (2002). *Discussant, Symposium on New Directions in Statistical Learning, International Conference on Infant Studies*, Toronto, Ont., Canada.

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

Jusczyk, P. W., & Derrah, C. (1987). Representation of speech sounds by young infants. *Developmental Psychology, 23*, 648–654.

Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge: MIT Press.

Marcus, G., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule-learning in seven-month-old infants. *Science, 283*, 77–80.

Maye, J. (2000). Learning speech sound categories on the basis of distributional information. Unpublished doctoral dissertation, University of Arizona.

Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101–B111.

McCarthy, J. L. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry, 12*, 373–418.

Mehler, J. (1981). The role of syllables in speech processing: Infant and adult data. *Philosophical Transactions of the Royal Society, 295*, 333–352.

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. II). New York: Wiley.

Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology, 63*, 176–185.

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science, 26*, 393–425.

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

Newport, E. L. (1981). Constraints on structure: Evidence from American Sign Language and language learning. In W. A. Collins (Ed.), *Aspects of the development of competence. Minnesota Symposium on Child Psychology* (Vol. 14). Hillsdale, NJ: Erlbaum.

Newport, E. L. (1982). Task-specificity in language learning? Evidence from speech perception and American Sign Language. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art*. New York: Cambridge University Press.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11–28.

Newport, E. L., & Aslin, R. N. (2000). Innately constrained learning: Blending old and new approaches to language acquisition. In S. C. Howell, S. A. Fish, & T. Keith-Lucas (Eds.), *Proceedings of the 24th annual Boston University conference on language development*. Somerville, MA: Cascadilla Press.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (submitted). Learning at a distance: II. Statistical learning of non-adjacent dependencies in a nonhuman primate.

Nygaard, L. C., & Pisoni, D. B. (1995). Speech perception: New directions in research and theory. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication*. E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception and cognition* (2nd ed.). San Diego: Academic Press.

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

Pinker, S. (1994). *The language instinct*. New York: Morrow.

Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 88–94.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*, 493–515.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172–196.

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

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.

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: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*, 101–105.

Savin, H., & Bever, T. G. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior, 9*, 295–302.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*, 1599–1603.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. Ferguson & D. Slobin (Eds.), *Studies of child language development*. New York: Holt, Rinehart & Winston.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science, 9*, 325–329.

Wertheimer, M. (1938). Laws of organization in perceptual forms. In W. Ellis (Ed. and Trans.), *A source book of Gestalt psychology* (pp. 71–88). London: Routledge & Kegan Paul. (Original work published 1923).

Wertheimer, M. (1944). Gestalt theory. *Social Research, 11*, 78–99.

Wexler, K., & Culicover, P. (1980). *Formal principles of language acquisition*. Cambridge: MIT Press.