PSYCHOLOGICAL SCIENCE

Research Article

Finding Words and Rules in a Speech Stream

Functional Differences Between Vowels and Consonants

Juan M. Toro,[1] Marina Nespor,[2] Jacques Mehler,[1,3] and Luca L. Bonatti[1,4]

[1]International School for Advanced Studies; [2]University of Ferrara; [3]Laboratoire de Science Cognitive et Psycholinguistique, CNRS and EHESS, Paris, France; and [4]Université de Nantes, Nantes Atlantique Universités, Labécd (EA 3259)

ABSTRACT—*We have proposed that consonants give cues primarily about the lexicon, whereas vowels carry cues about syntax. In a study supporting this hypothesis, we showed that when segmenting words from an artificial continuous stream, participants compute statistical relations over consonants, but not over vowels. In the study reported here, we tested the symmetrical hypothesis that when participants listen to words in a speech stream, they tend to exploit relations among vowels to extract generalizations, but tend to disregard the same relations among consonants. In our streams, participants could segment words on the basis of transitional probabilities in one tier and could extract a structural regularity in the other tier. Participants used consonants to extract words, but vowels to extract a structural generalization. They were unable to extract the same generalization using consonants, even when word segmentation was facilitated and the generalization made simpler. Our results suggest that different signal-driven computations prime lexical and grammatical processing.*

Humans learn and process their native language remarkably efficiently, despite its complexity. To achieve competence, in few years a learner must acquire tens of thousands of words and a formidable web of regularities, retrieving information about both the language's discrete units (notably, words) and its structural properties (grammar) from continuous speech. How humans solve these problems is a topic of intense research. For decades, linguists assumed that language acquisition requires preexisting knowledge. However, the speech signal does not reveal its grammar or lexicon in any obvious way. Even assuming powerful innate structures, one still needs to explain what mediates the relation between linguistic structures and the perceived signal. Recent research suggests that the signal is richer than previously thought, providing many statistical cues to be exploited. If learners were good statisticians, they could track the distribution of the basic elements in the signal and retrieve at least part of what needs to be learned without relying on preexisting structures.

Indeed, humans can compute statistical relations (in particular, transitional probabilities, or TPs) that can be useful to identify lexical elements in continuous speech (e.g., Saffran, Newport, & Aslin, 1996). This ability is not specifically linguistic, as it surfaces in several unrelated cognitive domains, such as image and tone processing (e.g., Fiser & Aslin, 2001, 2002); nor is it specifically human, as other mammals exhibit it (Newport, Hauser, Spaepen, & Aslin, 2004; Toro & Trobalòn, 2005).

Although TP computations are not specific to language processing, recent studies have provided evidence that the speech signal is richer in more language-specific dimensions, and that different computational mechanisms modulated by signal properties guide the search for distinct types of information. For example, TP computations, although useful for finding words in continuous speech, may not help in acquiring structural generalizations present in the same signal, whereas minimal signal differences (e.g., hidden segmentation cues or final-vowel lengthening) may trigger a fast, structure-sensitive mechanism that can grasp them (Endress & Bonatti, 2007; Peña, Bonatti, Nespor, & Mehler, 2002). If different mechanisms process speech, the signal itself may indicate a "division of labor" to the

---

Address correspondence to Luca L. Bonatti, Université de Nantes, Nantes Atlantique Universités, Labécd (EA 3259), BP 81227, 44312, Nantes Cedex 3, France, e-mail: lucabonatti@mac.com.

learner, such that a particular aspect of the signal may provide information for a particular mechanism and the complexity of the learning problems is reduced.

Prosody offers a good example. Prosodic information may help bootstrap the acquisition of syntax, even in the absence of lexical information, by signaling word order or syntactic phrase boundaries (Christophe, Nespor, Guasti, & Van Ooyen, 2003). So, whereas word extraction resorts to extensive stimulus sampling by means of general statistical abilities, structure extraction may be particularly sensitive to prosodic information useful in uncovering generalizations in speech.

Given that vowels are the main carriers of prosody, a radical possibility is that this division of labor is realized within words as a difference between information carried by vowels and information carried by consonants. Indeed, in natural languages, vowels and consonants tend to fulfill different roles. Lexical information is carried mostly by consonants, with Semitic languages representing the extreme case of word roots made up only of consonants. By contrast, vowels, which carry mainly prosodic information through vowel harmony, pitch changes, or lengthening, may signal aspects of syntactic structure (Nespor & Vogel, 1986), cueing the listener about how units, such as individual words, are organized. Even in languages in which the distinction is less sharp than in Semitic languages, vowels and consonants undergo several operations suggesting that they are categorically distinct and independently represented in separate tiers (Goldsmith, 1976). We (Nespor, Peña, & Mehler, 2003) have proposed the general hypothesis that vowels and consonants serve partially different roles, with vowels encoding more information about syntactic structure, and consonants instead conveying information about lexical identity. We will call this the *CV hypothesis*. Artificial-language learning offers a good way to test it, because studying how participants treat completely unknown miniature languages may reveal their natural learning processes. Indeed, the roles of prosody and statistical computations can be effectively studied with artificial languages (Shukla, Nespor, & Mehler, 2007).

A first prediction of the CV hypothesis is that participants should rely more on consonants than on vowels to identify words in a continuous artificial stream. In a series of word-segmentation experiments, we confirmed this prediction (Bonatti, Peña, Nespor, & Mehler, 2005; Mehler, Peña, Nespor, & Bonatti, 2006). Participants segmented words by exploiting the statistical coherence among consonants, but failed to segment words when vowels carried the same statistical relations. Nazzi (2005) showed analogous asymmetries between consonants and vowels when toddlers learned new words.

A second, crucial prediction of the CV hypothesis remains untested. Just as consonants are the main focus of statistical computations in extracting words from a stream, vowels should play a special role in extracting generalizations. Thus, the CV hypothesis predicts a complete reversal in how participants rely on consonants or vowels, according to whether words or structural regularities are hidden in a speech stream. In the study reported in this article, we tested this prediction. To assess a functional dissociation between consonants and vowels, we synthesized monotonous streams of consonant-vowel (CV) syllables. In Experiment 1, the stream contained statistically coherent consonantal frames, whereas vowels followed a simple structural organization. In Experiment 2, we reversed the roles of consonants and vowels by creating a "mirror" stream in which statistical coherence was based on vowels and structural organization on consonants. If participants are sensitive to the difference in information encoded by consonants and vowels, they should segment words using consonants, but also find generalizations using the underlying structure carried by vowels in streams such as the one used in Experiment 1. In contrast, they should fail both tasks when they hear a stream such as the one used in Experiment 2.

EXPERIMENT 1

Participants

Participants were 15 Italian students (8 females, 7 males; mean age = 22.5 years, range = 19–28). They received $10 for their participation.

Stimuli and Procedure

We created a continuous stream of CV syllables by concatenating 12 CVCVCV nonsense words, each of which had one of two consonant sequences and one of six vowel sequences. The vowel sequences all exhibited an ABA pattern; that is, the first and last vowels were identical (see Table 1). Syllable frequency was equalized across the stream. Between words, we inserted sequences of the same syllables composing words, in order to avoid immediate repetitions and simple alternations of consonantal frames. These *noise syllables* were arranged in one- to

TABLE 1
Experimental Stimuli in Experiment 1

	Segmentation test		Generalization test	
Familiarization words	Words	Part-words	Rule-words	Nonrule-words
tapena	tepane	penabe	biduki	buduki
tapona	tapona	dokato	budiku	biduku
tepane	topano	paneba	tipuni	tupuni
tepona	topeno	dakoba	tupinu	tipunu
topano	badeka	naboda		
topeno	bedake	katepa		
badeka	bedoke	nebade		
badoka	bodeko	kobado		
bedake				
bedoke				
bodako				
bodeko				

three-syllable chunks and randomly distributed across the stream. They constituted a total of 30% of the stream. Within each word, only two different syllables could immediately follow any given syllable. However, because of the presence of these "noise" syllables, within-word syllable TPs were .33 rather than .5. For the same reason, although only one consonant could follow any given consonant within a word, within-word consonant TPs were .7 rather than 1. The within-word vowel TP was .4. Between words, syllable TPs ranged from .027 to .077, vowel TPs ranged from .16 to .4, and consonant TPs were .16. We synthesized the stream with the MBROLA Italian database IT4 (Dutoit, Pagel, Bataille, & Vreken, 1996), setting syllable F0 to 240 Hz and phoneme duration to 120 ms.

For the test, we synthesized four types of trisyllabic items: *words*, *part-words*, *rule-words*, and *nonrule-words* (see Table 1). Words were composed as explained in the previous paragraph. Part-words were syllable sequences straddling word boundaries, but respecting the ABA vocalic structure. Rule-words retained both the consonant sequences in the words and the ABA vowel structure, but their vowels never appeared in the familiarization stream. Nonrule-words were like rule-words, but with either AAB or ABB vowel structure, in equal proportions. Thus, only the TPs between consonants could separate words from part-words, which had lower consonant TPs because they straddled word boundaries. And only compliance to the ABA vowel structure could distinguish rule-words, which never appeared in the stream but instantiated it, from nonrule-words, whose syllables were identical to those of rule-words but violated the ABA structure.

Participants were tested individually in a silent room. An Apple PowerPC running PsyScope X (Cohen, MacWhinney, Flatt, & Provost, 1993)[1] controlled the experiment, and the stimuli were presented through headphones. After 10 min of familiarization, participants were given an auditory two-alternative forced-choice test, with two kinds of test pairs: words versus part-words (*segmentation test*) and rule-words versus nonrule-words (*generalization test*). Participants were asked to consider how likely they thought it was that the members of each pair belonged to the familiarization language and to indicate the item in which they had greater confidence. The segmentation test assessed the ability to segment the stream by relying on TP computations among consonants, and the generalization test assessed the ability to extract a generalization by relying on vowel structure. The 16 test trials (8 for each comparison) were interleaved in semirandom order, with the restriction that no more than two trials of the same type could occur successively. To balance the generalization-test items with the segmentation-test items, we repeated each pair in the generalization test, reversing the order of the items in the generalization pairs. In each trial, test items were separated by a 500-ms pause.

Results and Discussion

Participants preferred words to part-words in the segmentation test ($M = 63.3\% \pm 8.7$), $t(14) = 5.87$, $p_{rep} = .96$, $d = 2.16$, and rule-words to nonrule-words in the generalization test ($M = 66.6\% \pm 11.2$), $t(14) = 5.73$, $p_{rep} = .96$, $d = 2.09$ (see Fig. 1). Performance on the two tests did not differ, $t(14) = 0.84$, $p_{rep} = .56$, $d = 0.32$. Thus, listeners can readily use distributional information about consonants to find words in a continuous speech stream; this finding confirms the role of consonants in word identification (Bonatti et al., 2005; Newport & Aslin, 2004). More important, the experiment shows that listeners can also extract a structural regularity from the vowels in a speech stream and apply this regularity to novel items productively. Thus, over a single stream, statistical information and structural information are both extracted, but used for different purposes—the former to identify words and the latter to generalize about structure—and as predicted by the CV hypothesis, consonants are mainly used to obtain statistical information, and vowels to obtain structural information.

To assess if any preference for particular test items affected the results, we ran a control experiment, familiarizing participants ($N = 15$) with a stream composed of the same syllables as in Experiment 1, but presented in random order, and then testing participants with the same items as in Experiment 1. There was no preference in either the segmentation test ($M = 48.3\% \pm 16.9$), $t(14) = 0.38$, $p_{rep} = .34$, $d = 0.14$, or the generalization test ($M = 45.8\% \pm 18.7$), $t(14) = 0.86$, $p_{rep} = .56$, $d = 0.31$. These results differed from those of Experiment 1, $t(28) = 3.04$, $p_{rep} = .95$, $d = 1.11$, and $t(28) = 3.69$, $p_{rep} = .95$, $d = 1.31$, respectively. Thus, the pattern of results in Experiment 1 was indeed due to the statistical- and rulelike properties of the familiarization stream.

It is worth emphasizing that the two computations were performed on different basic units. Previous studies showed that participants need segmentation cues to extract structural generalizations from a stream (Endress & Bonatti, 2007; Peña et al., 2002). Because the stream in Experiment 1 was continuous, it may be argued that participants' success at extracting generalizations from vowels contradicts our previous results. However, in our previous experiments (Peña et al., 2002), participants had to exploit the same representations—syllable sequences—to extract both words and rules. In contrast, in the present Experiment 1, they could use consonant sequences to extract words and vowels to extract syntaxlike regularities. Having statistical dependencies and abstract regularities on separate levels of representation (the consonantal and the vocalic tiers, respectively) may enhance the efficiency of both computations, provided that the underlying representations can support such computations.

The CV hypothesis offers a stronger prediction. If it is true that consonants identify word roots and vowels carry structural information, participants should fail both the segmentation and the generalization tests if the roles of consonants and vowels are reversed in the signal. Alternatively, if statistical computations

---

[1] The PsyScope code, rewritten and adapted to run under Mac Os X by our laboratory, is freely available at http://psy.ck.sissa.it/.

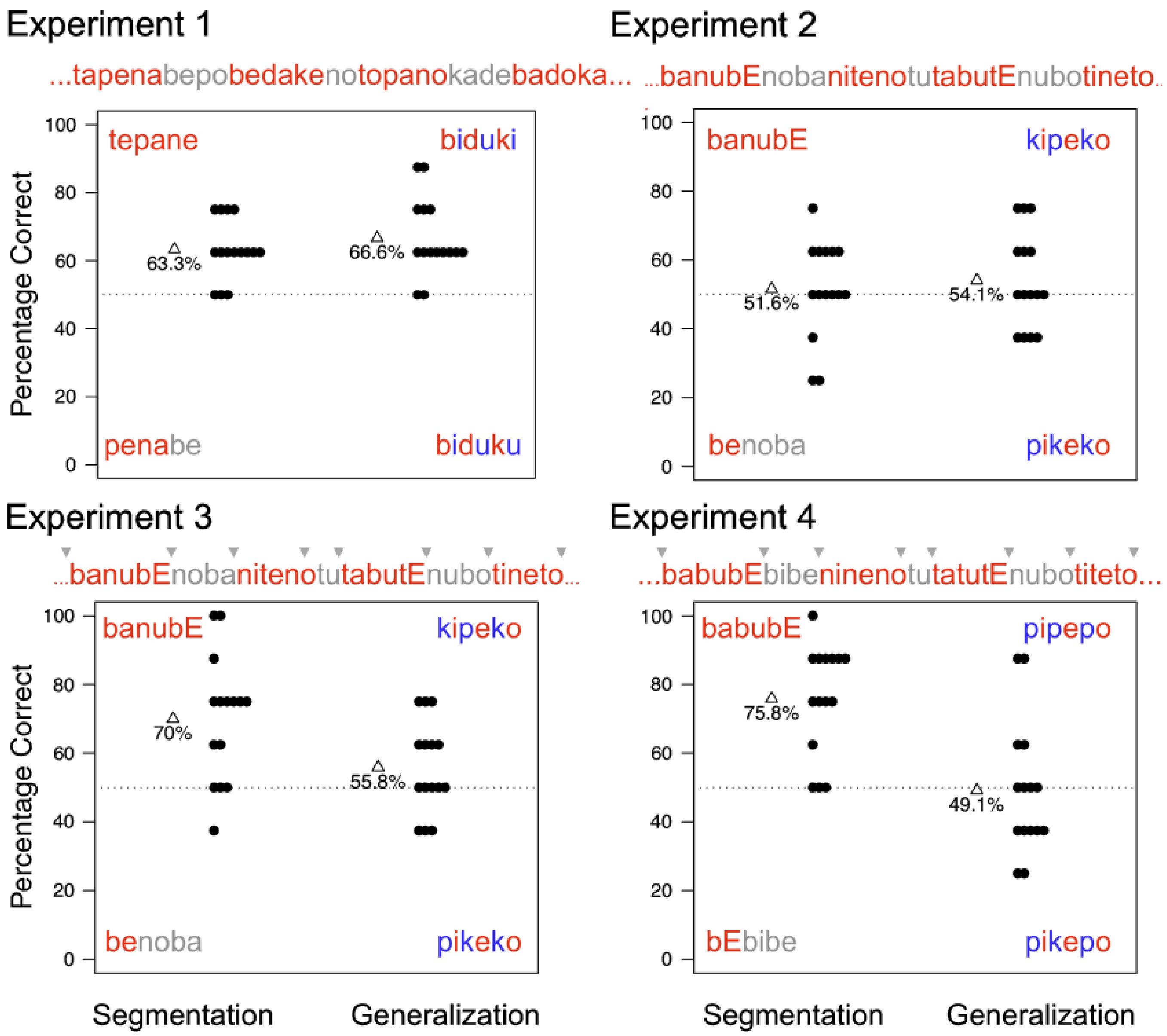

Fig. 1. Group means (open triangles) and individual scores (black circles) in the segmentation and generalization tests in Experiments 1 through 4. A partial example of the familiarization stream used is displayed over each panel; words are shown in red, noise syllables are in gray, and triangles signal pauses between syllables. Each graph shows examples of test items (words and rule-words above the data points, part-words and nonrule-words below); phonemes that did not change with respect to familiarization are in red, the blue letters are novel phonemes inserted only in the test items, and noise syllables (used to compose part-words) are in gray.

and structural generalizations can each be performed equally efficiently over vowels and over consonants, participants should succeed at both tests even when the roles of consonants and vowels are reversed. We tested these alternative predictions in Experiment 2.

EXPERIMENT 2

Participants, Stimuli, and Procedure

Participants were 15 Italian students (10 females, 5 males; mean age = 21.7 years, range = 19–28). They received $10 for their participation.

We synthesized a familiarization stream with statistical and structural properties identical to those of Experiment 1, but the roles of consonants and vowels were reversed such that the vowel TPs in Experiment 1 matched the consonant TPs in Experiment 2 and vice versa. Thus, the vowel tier implemented word frames, and the consonant tier implemented the ABA generalization. Twelve artificial words, each with CVCVCV structure, were created. Each word contained one of two vowel sequences and one of six consonant sequences instantiating an ABA pattern (see Table 2). The artificial words were concatenated as in Experiment 1. Test items were also created as in Experiment 1, but using novel consonants to construct rule-words and nonrule-

TABLE 2

Experimental Stimuli in Experiments 2 and 3

	Segmentation test		Generalization test	
Familiarization words	Words	Part-words	Rule-words	Nonrule-words
banubE	banubE	benoba	kapukE	kakupE
batubE	nabunE	butEbi	pakupE	pakukE
nabunE	natunE	nEbine	kipeko	kikepo
natunE	tanutE	netona	pikepo	pikeko
tabutE	binebo	nobanu		
tanutE	bitebo	nutEni		
binebo	nibeno	tEnite		
bitebo	tineto	tonatu		
nibeno				
niteno				
tibeto				
tineto				

words. Thus, vowels were crucial to identifying words, and consonants were the basis for finding the generalization. The procedure was identical to that of Experiment 1.

Results and Discussion

Participants expressed no preference between words and part-words ($M = 51.6\% \pm 14.1$), $t(14) = 0.45$, $p_{\text{rep}} = .39$, $d = 0.16$, or between rule-words and nonrule-words ($M = 54.1\% \pm 13.9$), $t(14) = 1.16$, $p_{\text{rep}} = .67$, $d = 0.51$ (see Fig. 1). Both results differed from those of Experiment 1, $t(28) = 2.72$, $p_{\text{rep}} = .92$, $d = 0.99$, for the segmentation test and $t(28) = 2.70$, $p_{\text{rep}} = .92$, $d = 0.96$, for the generalization test. These results document a sharp asymmetry in the roles that consonants and vowels play in on-line speech processing. Not only were participants unable to use the distribution of vowels to segment words from a continuous stream, but they were also unable to generalize the structural organization of consonants within a word. These data suggest that there are different mechanisms that perform different computations on a speech stream and extract different types of information from it.

In a control experiment, participants ($N = 15$) were familiarized with a stream composed of the same syllables as in Experiment 2, but randomly scrambled, and were tested with the same test items. These participants showed no preference either in the segmentation test ($M = 55\% \pm 16.9$), $t(14) = 1.14$, $p_{\text{rep}} = .66$, $d = 0.41$, or in the generalization test ($M = 51.6\% \pm 11.4$), $t(14) = 0.56$, $p_{\text{rep}} = .44$, $d = 0.19$, just as in Experiment 2, $t(28) = 0.58$, $p_{\text{rep}} = .45$, $d = 0.21$, and $t(28) = 0.53$, $p_{\text{rep}} = .43$, $d = 0.19$, respectively. Thus, no prior preference for test items affected the results, but most important, familiarization in Experiment 2 did not have a better effect than exposure to a random stream of syllables. This result reinforces our conclusion that participants could not use the statistical structure of vowels to extract words or the rulelike structure of consonants to extract structural information.

The results of Experiment 2 may have an alternative explanation. Perhaps participants failed the generalization test not because they found it difficult to generalize over consonants, but because they failed to segment the stream. If participants did not segment the words, they could not have extracted the structural organization of consonants, which was defined over words. To explore this possibility, in Experiment 3 we used a technique that bypasses the need to compute statistical relations to segment words. In previous work (Peña et al., 2002), we showed that 25-ms pauses at word edges greatly enhance segmentation, thereby also enhancing the ability to capture generalizations. Thus, Experiment 3 used the same familiarization stream as in Experiment 2, but 25-ms pauses were inserted at word boundaries. After participants had segmented words, if they could use consonants to project structural generalizations, they would be able to succeed on the generalization test. If, instead, failure to generalize to structure from consonants in Experiment 2 was due to speech-processing mechanisms that disregard consonants in extracting structure, participants in Experiment 3 would still fail to generalize despite passing the segmentation test.

EXPERIMENT 3

Participants, Stimuli, and Procedure

Participants were 15 Italian students (8 females, 7 males; mean age = 23.3 years, range = 21–29) who received $10 for their participation. The familiarization stream was the same as in Experiment 2, except for 25-ms pauses introduced at word edges. The test items and procedure were identical to those in Experiment 2.

Results and Discussion

Participants preferred words to part-words ($M = 70\% \pm 18.1$), $t(14) = 4.26$, $p_{\text{rep}} = .98$, $d = 1.56$, but showed no preference between rule-words and nonrule-words ($M = 55.8\% \pm 13.2$), $t(14) = 1.7$, $p_{\text{rep}} = .8$, $d = 0.62$ (see Fig. 1). The difference between the tests was significant, $t(14) = 2.67$, $p_{\text{rep}} = 0.92$, $d = 0.89$. Although participants segmented words better than in Experiment 2, $t(28) = 3.08$, $p_{\text{rep}} = 0.95$, $d = 1.13$, their ability to generalize did not improve, $t(28) = 0.33$, $p_{\text{rep}} = 0.33$, $d = 0.1$. Thus, even when segmentation cues help identify words in the stream, participants still fail to project generalizations over consonants. The failure to generalize in Experiment 2 cannot be attributed to a failure in segmenting words from a stream.

It is intriguing that participants could not generalize the simple ABA structure over consonants. After all, 7-month-old infants master similar ABA structures (Marcus, Vijayan, Rao, & Vishton, 1999). In Experiment 4, we maximally simplified the consonantal sequences by imposing an AAA consonant structure. Participants needed only to notice the repetition of a single element to be able to generalize correctly.

TABLE 3
Experimental Stimuli in Experiment 4

Familiarization words	Segmentation test: Words	Segmentation test: Part-words	Generalization test: Rule-words	Generalization test: Nonrule-words
babubE	babubE	bEbibe	kakukE	kapukE
nanunE	nanunE	bubEbi	kikeko	kipeko
tatutE	tatutE	nenona	papupE	pakupE
bibebo	bibebo	nEnine	pipepo	pikepo
nineno	nineno	totatu		
titeto	titeto	tutEti		

EXPERIMENT 4

Participants, Stimuli, and Procedure

Participants were 15 Italian students (11 females, 4 males; mean age = 22.4 years, range = 20–27) who received $10 for their participation. The familiarization stream used the same vocalic frames and consonants as in Experiments 2 and 3; however, a single consonant was repeated within each word (total of six words; see Table 3). Noise syllables and 25-ms pauses separated words, as in Experiment 3. Test items were also like those in Experiments 2 and 3, except for the repetition of the consonants in words and rule-words. Because rule-words had AAA consonant structure, we used nonrule-words with ABA consonant structure. This choice allowed us to control another aspect of Experiments 2 and 3. If participants had failed to generalize because they disfavored an ABA consonant structure, transforming the ABA items into nonrule-words would maximize the possibility of obtaining a positive result in Experiment 4. As in Experiments 2 and 3, vocalic information was irrelevant in the generalization test; participants needed to notice only nonidentity between consonants to differentiate invalid items from rule-abiding ones. The procedure was identical to that of Experiment 3.

Results and Discussion

Participants preferred words to part-words ($M = 75.8\% \pm 15.9$), $t(14) = 4.26$, $p_{\text{rep}} = .98$, $d = 2.29$, but still failed to prefer either rule-words or nonrule-words ($M = 49.1\% \pm 19.1$), $t(14) = 0.16$, $p_{\text{rep}} = .22$, $d = 0.06$ (see Fig. 1). Because the stream was covertly segmented, success on the segmentation test was expected. However, despite the greater simplicity of the generalization test, participants could not generalize structural information over consonantal sequences, just as in Experiments 2 and 3. This result demonstrates that in striking contrast to generalizations over vowels, structural generalizations over consonants seem to elude learners of an unknown language. Together, these results strongly confirm the CV hypothesis, suggesting that different linguistic representations generated by vowels and consonants in a speech signal play different roles in language processing.

GENERAL DISCUSSION

For decades, linguists and cognitive scientists have argued that the stimulus is too poor to account for the rich knowledge underlying human language competence. Several recent discoveries have revealed that the signal is not so poor after all. Adults and infants are sensitive to nontrivial statistical relations in the input that can help them solve the many puzzles encountered in language learning. Because this sensitivity is not specific to language, researchers have raised the possibility that learners acquire language only by means of general-purpose mechanisms that track statistical relations over large samples (e.g., Bates & Elman, 1996; Elman et al., 1996).

If all language were processed by such general-purpose mechanisms, consonants and vowels would be equally valid representations for computing statistics or drawing generalizations over the underlying structure of words. Instead, we found a striking asymmetry in the roles of consonants and vowels in language processing. Using artificial speech streams in which words and rules were implemented over either consonants or vowels, we showed that participants identify words by computing statistical dependencies among consonants, but fail to identify words when vowels carry the same dependencies. Moreover, participants can extract simple structural generalizations when vowels instantiate them, but fail to extract generalizations instantiated by consonants. These results are predicted by the CV hypothesis, but remain unexplained curiosities if one assumes that learning occurs only via general-purpose statistical mechanisms.

In contrast to this general-purpose view of the role of statistical processes in learning, there is a more encompassing view according to which perceptual (i.e., extralinguistic) factors constrain a general-purpose statistical learning device (e.g., Creel, Newport, & Aslin, 2004; Newport & Aslin, 2004; Seidenberg, MacDonald, & Saffran, 2002) and the interaction between statistical computations and perceptual constraints leads to language acquisition. Our results challenge this view, too. Perceptual constraints may render vowels more salient than consonants, or vice versa, but it is difficult to see how they would lead consonants to assume a prominent role in word identification and vowels to assume a prominent role in syntax. Likewise, if some perceptual factor rendered vowels more salient than consonants for computing syntactic-like regularities, it is difficult to see how, at the same time, it would make vowels ineffective for word identification.

Can some distributional information extracted from linguistic experience explain the asymmetry we documented, without appealing to functional differences between vowels and consonants? Indeed, for any phenomenon that some researchers have attributed to structural differences in language, other researchers have argued that some statistics could produce the same results (e.g., Seidenberg et al., 2002). This is not surprising. Because every law is bound to generate statistically measurable regularities, it will always be possible to find a statistic that correlates with a structural phenomenon. After all, this is what laws do: produce lawlike instances. Therefore, the mere fact that some statistical measure correlates with a structural phenomenon is not informative.

The claim that a specific distributional regularity is responsible for a specific structural phenomenon is more substantial. Recently, it has been argued that information-theoretic measures, such as mutual information, can retrieve many linguistic structures (Keidel, Jenison, Kluender, & Seidenberg, 2007; Redington, Chater, & Finch, 1998; Swingley, 2005), as well as explain participants' reliance on consonants rather than vowels to identify words in a continuous speech stream. Thus, Keidel et al. (2007) reported data indicating that in French (the native language of the participants in our prior experiments), consonant sequence types outnumber vowel sequence types in CVCVCV words (the word structure we used in the present study and also in Bonatti et al., 2005), at a ratio of 1.45:1. When the distributions of vowel- and consonant-sequence tokens are considered, this imbalance translates into a 0.51-bit advantage in mutual information for French consonant sequences over vowel sequences. Thus, Keidel et al. concluded, speakers preferentially attend to consonants "because a lifetime of linguistic experience indicates that consonants are more informative than vowels" (p. 923).

However, the direction of causality is difficult to determine. French consonants carry more mutual information than vowels, but it is also part of the CV hypothesis that consonant sequences will be more diverse than vowel sequences—otherwise word roots could not be differentiated, and the lexicon would be impoverished. Thus, the CV hypothesis predicts that consonant sequences will be more numerous and will carry more information than vowel sequences (Bonatti, Peña, Nespor, & Mehler, 2007; Nespor et al., 2003). Does causation flow from distributional information to structure, or from structure to distributions in the world?

The current results make some progress in determining the direction of the flow. Unlike French, Italian has more consonants (21) than vowels (7). Distributionally, the role of consonants is even more important, with a 13:1 ratio of consonant to vowel sequences in three-syllable words of the form we used, and a mutual-information advantage of 1.2 bits for consonants.[2] Therefore, if participants' preferences were driven not by a functional difference between vowels and consonants, but by differences in mutual information built up through years of linguistic experience, our participants, who were native speakers of Italian, should have relied on consonants even more than native speakers of French (the participants in Bonatti et al., 2005), regardless of the task. However, our Italian participants did not use consonants to generalize even to the easiest structure instantiated by consonants, yet were able to extract the generalization when vowels carried the same structure. Thus, the present experiments indicate that it is unlikely that mutual information (or other distributional measures extracted from lifelong experience with a native language), or perceptual constraints together with statistical computations, can explain away the functional difference between vowels and consonants.

We emphasize that the CV hypothesis does not imply that statistics cannot be computed over vowels or that structural generalizations cannot be extracted over consonants. In fact, studies (Bonatti et al., 2005; Newport & Aslin, 2004) have shown that under redundant conditions (e.g., immediate repetitions of the same patterns), vowels can be used for word identification. Likewise, generalizations can be made over consonants, as demonstrated by many syntactic regularities in natural languages (Berent, Marcus, Shimron, & Gafos, 2002). However, our results show that within the same lexical item, representations built on vowel sequences and representations built on consonant sequences preferentially carry different types of information, which nicely correspond to the difference between lexicon and syntax. Our results suggest that the linguistic nature of the representations computed from the signal being processed, rather than some lower-level perceptual factor, constrains the domain of application of mechanisms dedicated to retrieve either lexical or syntactic information. Taking advantage of such different representations may provide the learner with important cues for acquiring lexical and syntactic knowledge.

Thus, as many researchers have argued, the signal is indeed very rich. However, our results show that its richness lies in the "eye" of the beholder. Because humans are endowed with a complex set of language-learning mechanisms dedicated to building different aspects of linguistic competence, vowels and consonants emerge from the signal as the main carriers of different kinds of information.

Acknowledgments—This research was supported by Human Frontiers Science Project Grant RGP 68/2002, Regione Friuli-Venezia-Giulia Grant 2005016300, and European Union CALACEI 12778. We thank R. Aslin, A. Endress, M. Shukla, and three anonymous reviewers for their very helpful comments, and M. Nathan for her impeccable editorial work.

REFERENCES

Bates, E., & Elman, J. (1996). Learning rediscovered. *Science*, *274*, 1849–1850.

---

[2] Because no large database exists for Italian, we analyzed the 3,150,075 entries of the CoLFIS (Corpus e Lessico di Frequenza dell'Italiano Scritto) corpus of written Italian (Laudanna, Thornton, Brown, Burani, & Marconi, 1995). CoLFIS does not code oral phonological forms, and the exact phonological structure of a word often cannot be predicted without accentual information (which CoLFIS does not code). Therefore, we approximated the ratio of consonant sequences to vowel sequences by using an automatic transformation of graphemes into phonemes, treating semivowels as consonants. Because /E/ cannot be distinguished from /e/ in written Italian, and /O/ cannot be distinguished from /o/, our transformation underestimated vowel sequences. However, /E/ and /O/ are fairly rare and nondistinctive phonemes, and the 13:1 ratio should not change significantly with a more precise transliteration.

Berent, I., Marcus, G.F., Shimron, J., & Gafos, A.I. (2002). The scope of linguistic generalizations: Evidence from Hebrew word formation. *Cognition*, *83*, 113–139.

Bonatti, L.L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, *16*, 451–459.

Bonatti, L.L., Peña, M., Nespor, M., & Mehler, J. (2007). On consonants, vowels, chickens, and eggs. *Psychological Science*, *18*, 924–925.

Christophe, A., Nespor, M., Guasti, M.T., & Van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: The case of the head-direction parameter. *Developmental Science*, *6*, 211–220.

Cohen, J.D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, *25*, 257–271.

Creel, S.C., Newport, E.L., & Aslin, R.N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1119–1130.

Dutoit, T., Pagel, V., Bataille, F., & Vreken, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. In *ICSLP 96: Proceedings, Fourth International Conference on Spoken Language Processing* (Vol. 3, pp. 1393–1396). New York: IEEE.

Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Endress, A., & Bonatti, L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition*, *105*, 247–299.

Fiser, J., & Aslin, R.N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, *12*, 499–504.

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.

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic Analysis*, *2*, 23–68.

Keidel, J.L., Jenison, R.L., Kluender, K.R., & Seidenberg, M.S. (2007). Does grammar constrain statistical learning? Commentary on Bonatti, Peña, Nespor, and Mehler (2005). *Psychological Science*, *18*, 922–923.

Laudanna, A., Thornton, A.M., Brown, G., Burani, C., & Marconi, L. (1995). Corpus e lessico di frequenza dell'Italiano scritto. Retrieved September 2006 from http://www.istc.cnr.it/material/database/colfis I

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

Mehler, J., Peña, M., Nespor, M., & Bonatti, L.L. (2006). The "soul" of language does not use statistics: Reflections on vowels and consonants. *Cortex*, *42*, 846–854.

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, *98*, 13–30.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, *ii*(2), 201–227.

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

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

Newport, E.L., Hauser, M.D., Spaepen, G., & Aslin, R.N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*, 85–117.

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

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

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

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

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, *54*, 1–32.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*, 86–132.

Toro, J.M., & Trobalòn, J.B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics*, *67*, 867–875.

(RECEIVED 3/16/07; REVISION ACCEPTED 7/9/07)