Commentary

Does Grammar Constrain Statistical Learning?

Commentary on Bonatti, Peña, Nespor, and Mehler (2005)

James L. Keidel, Rick L. Jenison, Keith R. Kluender, and Mark S. Seidenberg

University of Wisconsin–Madison

Many studies indicate that statistical learning plays an important role in language acquisition (Saffran & Sahni, in press). Bonatti, Peña, Nespor, and Mehler (2005) presented evidence that such learning is guided by innate grammatical knowledge, a potentially important discovery. Here we provide data suggesting that their results may instead have resulted from facts about French that can be learned from experience.

In the study by Bonatti et al. (2005), subjects learned an artificial language in which the words were strings of alternating consonants (C) and vowels (V): CVCVCV (e.g., *puragi*). In one condition, transition probabilities between consonants within a word were 1.0, but vowels varied: For example, the consonants *p_R_g_* appeared only in that order, but each blank could be filled by multiple vowels. In a second condition, within-word transition probabilities between vowels were 1.0, and consonants varied (e.g., *_u_e_ã*). The key finding was that subjects picked up on the statistical regularity for consonants, but not vowels.

LANGUAGE MODULE REQUIRED?

According to Bonatti et al. (2005), this finding cannot be explained by domain-general statistical learning mechanisms: The vowel and consonant triples they used occur with similar frequencies, and French displays an atypical balance between consonants (17) and vowels (15).[1] Thus, "no prior numerical asymmetry could lead French listeners to adopt a strategy favoring [consonants over vowels]" (Bonatti et al., p. 457). Rather, Bonatti et al. argued, the language module encodes that consonants primarily mediate word identification, and vowels indicate grammatical functions. Thus, given a word-learning task, the language module directs the statistical processor to attend to the consonants. Moreover, if vowels are more perceptually salient than consonants, this property would bias an unconstrained statistical learner toward vowels. On this basis, one would predict the result opposite the one obtained. In fact, tamarin monkeys in a similar experiment learned transition probabilities of vowels, but not transition probabilities of consonants (Newport, Hauser, Spaepen, & Aslin, 2004). Hence, "only when a language module exists do consonants and vowels reverse their natural order of saliency" (Bonatti et al., p. 458).

Bonatti et al. were correct that consonants are more informative about word identity in French than vowels are, but this fact can be learned from experience. The 4,943 French CVCVCV words (based on a search of *Lexique 3*; New, Pallier, Ferrand, & Matos, 2001) contain 820 unique three-consonant tiers and 562 unique three-vowel tiers. Thus, on average, each consonant tier appears in 6.03 words, whereas each vowel tier appears in 8.8 words.

This effect can also be quantified using information theory (Shannon, 1948). The entropy (H) of a discrete random variable X in bits is defined as follows:

$$H(X) = -\sum_i p(x_i) \log_2 p(x_i), \qquad (1)$$

where $p(x_i)$ represents the probability of the ith element of the distribution. Given the 4,943 French CVCVCV words and their frequencies, the number of bits needed to encode a randomly selected word is 9.14. The mutual information (I) of tiers and words is as follows:

$$I(\text{words, tiers}) = H(\text{words}) - H(\text{words}|\text{tiers}) \qquad (2)$$

That is, the information in the tier-word mapping is equivalent to the entropy reduction created by knowing the vowel or consonant tier of a given lexical item. H(words) is unconditional and therefore independent of tier structure, so any increase in mutual information must be due to the degree to which the entropy of the words is reduced by knowing the tier. If all words could be uniquely identified by tier, then H(wordsltiers) would equal zero.

Address correspondence to Mark S. Seidenberg, University of Wisconsin–Madison, Department of Psychology, 1202 W. Johnson St., Madison, WI 53706, e-mail: seidenberg@wisc.edu.

[1] Bonatti et al. reported 16 vowels for French; *Lexique 3* uses 15, as the difference between closed and open /a/ is disappearing (Peereman & Dufour, 2003).

CONSONANTS ARE STATISTICALLY MORE INFORMATIVE THAN VOWELS

Equation 2 yields values of 7.14 for *I*(words, consonants) and 6.63 for *I*(words, vowels). We assessed the significance of this 0.51 difference by performing a bootstrap analysis to obtain a null sampling distribution for d. One thousand tier-word matrices with one value per word column were generated, because words have one consonant and one vowel tier. This analysis yielded an expected value for d of 0.26 ($SD = 0.04$). The 0.25 difference between the obtained (0.51) and bootstrapped (0.26) ds is highly statistically significant, indicating that consonants provide substantially more information about lexical identity than vowels.[2] Specifically, a 0.51-bit increase in mutual information (which includes the 0.26 bias) corresponds to an average reduction in uncertainty of .702 ($2^{-0.51}$) for a consonant tier over a vowel tier. The average improvement in identification is thus 1.42 (1/0.702), or about 40%, given no additional information.

Exposed to a stream of CV syllables, then, adult native French speakers may preferentially attend to consonants not out of any innate predisposition, but because a lifetime of linguistic experience indicates that consonants are more informative than vowels. According to this view, the tamarins' behavior differs from people's because only the latter have had years of exposure to French.

In summary, our analyses do not show that the conclusions of Bonatti et al. are necessarily incorrect, but rather show that there is an alternative interpretation consistent with their data. Additional research is needed in order to determine which interpretation is correct.

The study by Bonatti et al. (2005) is the most recent attempt to establish limits on the role of statistical learning in language acquisition (e.g., Marcus, Vijayan, Rao, & Vishton, 1999; Peña, Bonatti, Nespor, & Mehler, 2002). In each case, other researchers have subsequently identified other stimulus properties that could explain the results (e.g., Onnis, Monaghan, Richmond, & Chater, 2005; Seidenberg & Elman, 1999; Seidenberg, MacDonald, & Saffran, 2002). Two limitations of statistical-learning research have contributed to this pattern. First, most studies have focused on transition probabilities between phonemes or syllables; however, languages exhibit regularities involving other linguistic units that need to be taken into account if any bounds on statistical learning are to be identified. Second, little research has been tied to explicit theories of learning capable of explaining which statistics people encode. This makes it easy to overlook relevant regularities. We agree with Bonatti et al. that language learning is constrained in important ways, but whether the source of such constraints is innate grammatical knowledge, domain-general learning, or other perceptual or cognitive factors remains an open question.

REFERENCES

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.

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

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). A lexical database for contemporary French on internet: Lexique. *L'année Psychologique*, *101*, 447–462.

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.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, *53*, 225–237.

Peereman, R., & Dufour, S. (2003). Un correctif aux notations phonétiques de la base de données Lexique. *L'année Psychologique*, *103*, 103–108.

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

Saffran, J.R., & Sahni, S.D. (in press). Language development in infancy. In M. Spivey, M. Joanisse, & K. McRae (Eds.), *The Cambridge handbook of psycholinguistics*. New York: Cambridge University Press.

Seidenberg, M.S., & Elman, J.L. (1999). Do infants learn grammar with algebra or statistics? *Science*, *284*, 434–435.

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

Shannon, C.E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423.

(Received 7/11/06; Accepted 9/14/06; Final materials received 10/2/06)

[2]The nonzero bias in this calculation stems from the greater number of attested consonant tiers; the tier-word matrix for consonants has 820 rows, whereas the vowel-based matrix has 562 rows. The bootstrapped null sampling distribution for d is well approximated by a Gaussian distribution.