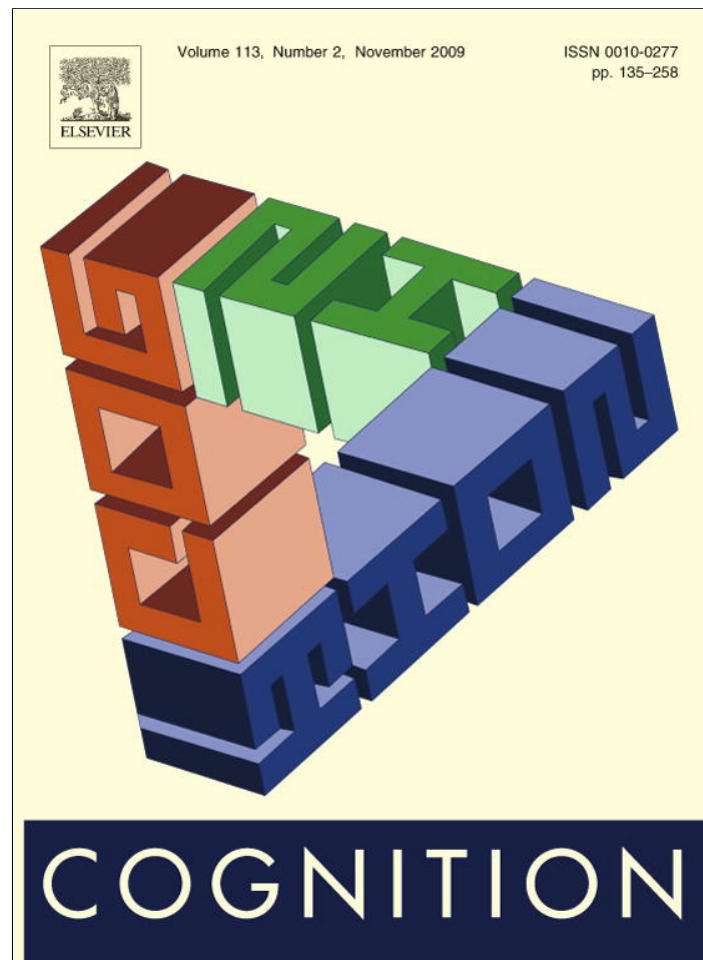


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Brief article

Learning in reverse: Eight-month-old infants track backward transitional probabilities

Bruna Pelucchi^{a,b,*}, Jessica F. Hay^a, Jenny R. Saffran^a

^a Department of Psychology and Waisman Center, University of Wisconsin – Madison, 1550 Highland Av, Madison, WI 53705, United States

^b Department of Evolutionary Biology, University of Ferrara, Italy

ARTICLE INFO

Article history:

Received 17 April 2009

Revised 30 July 2009

Accepted 30 July 2009

Keywords:

Infant language acquisition

Statistical learning

Backward transitional probability

ABSTRACT

Numerous recent studies suggest that human learners, including both infants and adults, readily track sequential statistics computed between adjacent elements. One such statistic, transitional probability, is typically calculated as the likelihood that one element predicts another. However, little is known about whether listeners are sensitive to the directionality of this computation. To address this issue, we tested 8-month-old infants in a word segmentation task, using fluent speech drawn from an unfamiliar natural language. Critically, test items were distinguished solely by their backward transitional probabilities. The results provide the first evidence that infants track backward statistics in fluent speech.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Language comprehension relies on the identification of basic lexical units. Preverbal infants quickly begin to recognize word-like sequences presented in fluent speech (Jusczyk & Aslin, 1995), suggesting that there must be some set of reliable cues marking word boundaries in the speech signal. Indeed, infants are sensitive to numerous acoustic and phonological properties correlated with word boundaries (for a recent review, see Saffran, Werker, & Werner, 2006).

In particular, infants' sensitivity to transitional probability (TP), the probability of event Y given event X , has received a great deal of attention (e.g., Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999). TP is typically calculated according to Eq. (1):

$$TP = P(Y|X) = \frac{\text{frequency}(XY)}{\text{frequency}(X)} \quad (1)$$

On this construal, the frequency of the first element in the pair, X , is normalized as a function of its overall frequency in the corpus. TP is thus a measure of the strength with which X predicts Y . Analyses of infant-directed speech corpora suggest that TP cues could, in principle, help infants find word boundaries, at least when acting in concert with other cues (Swingley, 2005). Furthermore, infants track TPs when exposed to a synthetic speech stream drawn from a miniature artificial language, in which TP is the only available cue to word boundaries (Aslin et al., 1998; Graf Estes, Evans, Alibali, & Saffran, 2007). Infants can also track TPs in natural speech drawn from an unfamiliar language, Italian (Pelucchi, Hay, & Saffran, 2009).

Despite this emerging body of work, remarkably little is known about the computational underpinnings of these findings. In particular, it is not clear whether infants compute TPs solely in a forward direction. In the original study by Aslin et al. (1998), the authors noted that while they were focused on forward TPs, there were a number of other pair-wise statistics that could be equally informative, in principle, because they normalized the frequency of co-occurrence of the pair by the frequencies of one or both of the individual elements. From a computational perspective, some sequential statistics are direction-independent, such as mutual information (e.g., Charniak, 1993; Swingley,

* Corresponding author. Address: Department of Psychology and Waisman Center, University of Wisconsin – Madison, 1550 Highland Av, Madison, WI 53705, United States.

E-mail address: pelucchi@wisc.edu (B. Pelucchi).

1999). Other statistics contain directional information. For example, TP can be defined in two different and symmetrical ways, depending on how the normalization step is performed. In addition to forward TPs (hereafter FTP) described in Eq. (1), it is also possible to compute backward TPs (hereafter BTP), as shown in Eq. (2), measuring the likelihood of X preceding Y:

$$\text{BTP} = P(X|Y) = \frac{\text{frequency}(XY)}{\text{frequency}(Y)} \quad (2)$$

Indeed, in a corpus analysis of English infant-directed speech, Swingley (1999) demonstrated that FTP and BTP are equally informative as independent cues to word boundaries.

The majority of the studies investigating TP computations have manipulated FTP, often not controlling for effects of BTP. These conditional probability statistics are typically correlated, making it difficult to assess the independent roles of FTPs and BTPs. Although the two measures are often identical in studies that use artificial language materials (e.g., Aslin et al., 1998; Graf Estes et al., 2007), they are likely to differ substantially in natural languages. In the one infant study in which these two cues were decoupled, Pelucchi et al. (2009) controlled for BTPs in their test items, leaving FTPs as the only useful cues (see Table 1). No study to date has tested whether infants can exploit BTP cues in the absence of informative FTP cues, though results from recent adult studies suggest that such learning is indeed feasible both in the auditory (Perruchet & Desauty, 2008) and visual domains (Jones & Pashler, 2007).

The aim of the current study was to determine whether infants track BTPs in fluent speech. Eight-month-old infants listened to a corpus of natural speech from an unfamiliar language. Unlike earlier studies of statistical learning using highly artificial stimuli, these materials have the complexity of speech found in infants' natural environments, increasing the ecological validity of the task (Pelucchi et al., 2009). After familiarization, infants were tested on two different types of familiar words: *High Transitional Probability words* (HTP-words, BTP = 1.0) versus *Low Transitional Probability words* (LTP-words, BTP = 0.33). Importantly, both word types occurred equally often during familiarization and shared the same trochaic stress pattern and the same FTP (1.0). Successful discrimination between HTP- and LTP-words would suggest that infants track BTPs while listening to natural speech.

2. Method

2.1. Participants

Thirty-two infants (mean age 8.4 months, range 8.0–9.0) were assigned to one of two counterbalanced conditions: Language A or B. All infants were monolingual English learners with no history of hearing or vision impairments; none had prior exposure to Italian or Spanish. Eight additional infants were excluded because of fussiness (6) or failure to attend (2).

2.2. Materials

Two counterbalanced languages were used to control for arbitrary listening preferences at test. Languages A and B both consisted of 12 grammatically correct meaningful Italian sentences. The four target words (*fuga*, *melo*, *bici*, and *casa*) each appeared six times in each corpus. These target words all followed a strong/weak stress pattern and were phonetically and phonotactically legal in English.

The critical manipulation concerned the BTPs between the syllables of the target words. In Language A, the syllables *fu*, *ga*, *me*, and *lo*, appeared only in the words *fuga* and *melo*. Consequently, the BTP of these two words was 1.0 (HTP-words). However, for *casa* and *bici* there were 12 additional occurrences of the syllables *sa* and *ci* in the corpus, always in weak (unstressed) position consistent with their stress level in the target words. As a consequence, the BTPs of *casa* and *bici* were 0.33 (LTP-words) relative to the Language A familiarization sentences.

Importantly, since the first syllables, *bi* and *ca*, only occurred in the context of the target words, the FTPs for *casa* and *bici* were 1.0. This is identical to the FTPs for the two HTP-words, *fuga* and *melo*. Thus, the only difference between HTP- and LTP-words is their BTPs: 1.0 for HTP-words, and 0.33 for LTP-words. In the counterbalanced Language B, HTP- and LTP-words were switched.

Each block of sentences was presented three times, for a total of 3 min. The resulting corpora for Languages A and B included 18 repetitions of each of the four target words, and 36 additional occurrences of the second syllables in the two LTP-words. Test items consisted of the 2 HTP-words and 2 LTP-words. A female native Italian speaker, naïve to the purpose of the experiment, recorded the stimuli in an infant-directed register. Test words were read in citation form and were digitally matched for length and amplitude, while preserving their original pitches.

Table 1

Comparison between FTP and BTP of the words and part-words from Saffran et al. (1996) and Perruchet and Desauty (2008), and for the HTP- and LTP-words from Pelucchi et al. (2009) and those used in the current experiment.

	Transitional probability							
	Saffran et al. (1996)		Perruchet and Desauty (2008) ^b		Pelucchi et al. (2009)		Current experiment	
	Words	Part-words ^a	Words	Part-words	HTP	LTP	HTP	LTP
FTP	1.0	.35	.20	.33	1.0	.33	1.0	1.0
BTP	1.0	.35	1.0	.20	1.0	1.0	1.0	.33

^a TPs at word boundaries are an average (.17/.48).

^b Values are from Experiment 1, backward group.

2.3. Procedure

Infants were tested using the Head Turn Preference Procedure as adapted by Saffran et al. (1996). During familiarization, Language A or B played from speakers mounted beneath two sidelights. The lights flashed contingent on looking behavior while the familiarization materials played continuously. Immediately after familiarization, 12 test trials were presented. Each trial consisted of a single test item repeated as long as the infant maintained a head-turn in the direction of the flashing light above the loudspeaker presenting the sound. All infants heard the same test items regardless of familiarization condition. Each of the 2 *HTP-word* and 2 *LTP-word* trials occurred three times, randomized by block. Test items that were *HTP-words* in Language A corresponded to *LTP-words* in Language B, and vice versa.

3. Results

As there were no significant differences in difference scores for the two counterbalanced languages [$t(30) = .30$, $p = .77$], the two conditions were combined in the subsequent analysis. A paired *t*-test revealed a significant difference in average looking time for *HTP-words* (10.06 s) versus *LTP-words* (8.91 s): $t(31) = 3.05$, $p < .01$ (see Fig. 1).

Since test items were matched for their trochaic pattern, frequency, and FTP, infants' successful test discrimination suggests sensitivity to BTP. The observed familiarity preference is consistent with results of Pelucchi et al. (2009), where FTP was manipulated using similar materials and procedures. Together, these results indicate that infants can keep track of TPs in fluent natural speech, computing probabilities of co-occurrence in both directions.

4. General discussion

Despite the burgeoning literature exploring the role of TP in word segmentation, remarkably little is known about the directional selectivity of this computation. Prior to the current study, data on infants' sensitivity to BTP was not available. These results provide the first evidence showing

that infants can successfully discriminate disyllabic sequences based on differences in their internal BTPs.

While intuition suggests that forward-going statistics should be more informative to learners, the acquisition of some types of linguistic structures would be facilitated by detection of backward-going statistics (e.g., Saffran, 2001, 2002; Saffran et al., 2008). For example, in languages that mark grammatical gender, backward TPs would help learners to discover which types of articles typically precede nouns; forward TPs are far less informative in this situation (Lew-Williams, personal communication). For example, in the Spanish article-noun sequence *la pelota*, *la* is not a good predictor of *pelota* since there are many possible different feminine nouns that could follow *la*. However, *pelota* is a good predictor of *la* since they are often paired. Even in English, BTPs are more useful than FTPs for discovering some relationships. Corpus analyses suggest that to discover the grammatical category 'noun', BTPs are far more informative than FTPs. For example, given the pair of words *the dog*, the backward probability (*the* preceding *dog*) is higher than the corresponding forward probability (*dog* following *the*; 0.25 vs. 0.002; Willits, Seidenberg, & Saffran, in press). In other languages, like Korean, determiners tend to follow nouns, and thus the forward probability is more informative than the corresponding backwards probability. Indeed, a recent adult study suggests that the degree of sensitivity to forward versus backwards probability may in fact be language dependent (Onnis, 2009). In general, forward and backward TPs, working together, likely help to constrain the set of potential alternatives across a range of learning problems.

The successful discrimination observed in this experiment has a further, potentially important, implication. To track BTPs in these materials, infants needed to compute probabilities over weak syllables. Unstressed syllables are less perceptually salient and thus potentially more elusive than strong syllables (Cutler & Foss, 1977; Tabossi, Burani, & Scott, 1995). In Italian, weak syllables are shorter, softer, and lower pitched than their corresponding strong syllables, which infants readily track in statistical learning tasks (Pelucchi et al., 2009). These results are thus important in that they demonstrate that infants can track the probabilities of these less prominent elements. Moreover, unlike earlier studies of statistical learning in infancy, the material used in this study contained the richness of natural speech. The results thus serve to emphasize the sophisticated abilities of infants who readily track the statistics of sounds in complex input despite differences in their acoustic realizations (e.g., coarticulation, nasalization, etc.).

The current studies illuminate the power of statistical learning mechanisms and extend our understanding of their computational range; infants readily track both forward and backward TP even in complex, natural linguistic input. These findings both constrain and challenge future studies that invoke statistical learning processes. Input statistics that appear to be difficult to learn based on weak forward-going statistics may be supported by strong backward-going statistics, and vice versa. Future research will need to pinpoint the means by which infant learners

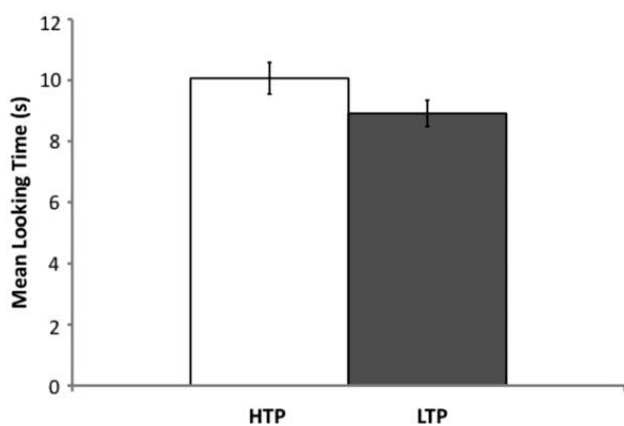


Fig. 1. Mean looking times (± 1 SE) to *HTP-words* and *LTP-words*.

integrate these sources of information, along with the myriad other cues that lurk amidst linguistic input.

Acknowledgements

This research was funded by grants from NICHD to JRS (R01HD37466) and JFH (F32-HD557032), a grant from the James F. McDonnell Foundation to JRS, and by a core grant to the Waisman Center (P30HD03352). We would like to thank the members of the Infant Learning Lab and especially Jessica Rich for her assistance in the conduct of this research, the families who generously contributed their time, and two anonymous reviewers for their helpful comments on a previous version of this manuscript.

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.
- Charniak, E. (1993). *Statistical language learning*. Cambridge, MA: MIT Press.
- Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech*, *20*, 1–10.
- Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science*, *18*, 254–260.
- Jones, J., & Pashler, H. (2007). Is the mind inherently forward thinking? Comparing prediction and retrodiction. *Psychonomic Bulletin and Review*, *14*(2), 295–300.
- Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1–23.
- Onnis, L. (2009). Statistical second language learning. Paper presented at Georgetown University Round Table, Washington, DC, March 13–15.
- Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674–685.
- Perruchet, P., & Desauty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory and Cognition*, *36*, 1299–1305.
- Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language*, *44*, 493–513.
- 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. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.
- Saffran, J. R., Hauser, M., Seibel, R. L., Kapfhammer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern learning by infants and cotton-top tamarin monkeys. *Cognition*, *107*, 479–500.
- Saffran, J., Johnson, E., Aslin, R., & Newport, E. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52.
- Saffran, J. R., Werker, J., & Werner, L. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In R. Siegler & D. Kuhn (Eds.), *Handbook of child development* (pp. 58–108). New York: Wiley.
- Swingle, D. (1999). Conditional probability and word discovery: A corpus analysis of speech to infant. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st annual conference of the cognitive science society* (pp. 724–729). Mahwah, NJ: Erlbaum.
- Swingle, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*, 86–132.
- Tabossi, P., Burani, C., & Scott, D. (1995). Word identification in fluent speech. *Journal of Memory and Language*, *34*, 440–467.
- Willits, J. A., Seidenberg, M. S., & Saffran, J. R. (2009). Verbs are lookING good in early language acquisition. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive science society* (pp. 2570–2575). Austin, TX: Cognitive Science Society.