Published in final edited form as:
Cognition. 2008 July ; 108(1): 271–280. doi:10.1016/j.cognition.2008.02.003.

The link between statistical segmentation and word learning in adults

Daniel Mirman[1], **James S. Magnuson**[1], **Katharine Graf Estes**[2], and **James A. Dixon**[1]

1*Department of Psychology, University of Connecticut and Haskins Laboratories*

2*Department of Psychology, University of California-Davis*

Abstract

Many studies have shown that listeners can segment words from running speech based on conditional probabilities of syllable transitions, suggesting that this statistical learning could be a foundational component of language learning. However, few studies have shown a direct link between statistical segmentation and word learning. We examined this possible link in adults by following a statistical segmentation exposure phase with an artificial lexicon learning phase. Participants were able to learn all novel object-label pairings, but pairings were learned faster when labels contained high probability (word-like) or non-occurring syllable transitions from the statistical segmentation phase than when they contained low probability (boundary-straddling) syllable transitions. This suggests that, for adults, labels inconsistent with expectations based on statistical learning are harder to learn than consistent or neutral labels. In contrast, infants seem learn consistent labels, but not inconsistent or neutral labels.

Keywords

statistical learning; word segmentation; word learning; language acquisition

The task of learning words from spoken input is an extremely difficult one in part because there are no consistent cues to word boundaries. Conditional probabilities of syllable sequences are one cue to word boundaries: within-word syllable sequences are much more likely than between-word syllable sequences. Saffran and colleagues reported that adults (Saffran, Newport, & Aslin, 1996) and infants (Saffran, Aslin, & Newport, 1996) can extract novel conditional probabilities from only a few minutes of exposure and use this statistical information for sequence segmentation. Saffran and colleagues found that both adults and infants could distinguish syllable sequences that contained high probability syllable transitions ("words", sequences that consistently occurred) from those that contained low probability syllable transitions ("partwords", sequences straddling a "word" boundary, thus occurring only occasionally). This statistically-based learning and segmentation ability could form part of a mechanism that supports language acquisition. However, it is also possible that, although adults and infants can extract these statistics in an explicit laboratory task, this learning has no

Correspondence: Daniel Mirman, Department of Psychology, University of Connecticut, 406 Babbidge Rd., Unit 1020, Storrs, CT 06269-1020, Phone: (860) 486-5722, Fax: (860) 486-2760, Email: daniel.mirman@uconn.edu.

connection to the mechanisms involved in learning new words. Only recently have researchers sought to demonstrate a direct link between statistical word segmentation and word learning.

A recent study tested the relation between statistical segmentation and object label learning in infants (Graf Estes, Evans, Alibali, & Saffran, 2007). Each infant was first exposed to a non-segmented syllable stream as in typical statistical segmentation studies. After this exposure phase, the infants completed a habituation-based object label learning phase. The infants were habituated to two label-object pairings, followed by two types of test trials: trials in which the original label-object pairings remained the same versus trials in which the pairings were switched. The difference in looking times between same and switch trials reflects the extent to which infants associated a particular object with a particular label; in other words, the extent to which infants learned object labels (see also Stager & Werker, 1997; Werker et al., 1998). For some of the infants the object labels contained high probability syllable transitions ("words") from the segmentation stream. For other infants the object labels contained low or zero probability syllable transitions ("partwords" or "nonwords", respectively). Graf Estes et al. found that when the object labels were "words", there was a looking time difference between same and switch trials, but there was no difference when the object labels were "partwords" or "nonwords" (labels containing syllable transitions that did not occur in the exposure stream). That is, infants learned object labels when those labels were consistent with the statistics of the preceding passive exposure phase, but not when the labels were inconsistent with those statistics. One interpretation of these results is that statistical segmentation facilitates word learning by creating memory traces that can then be mapped to meanings (object labels).

In the present work, we extended this finding by examining the link between statistical segmentation and word learning in adults. First, this extension allowed us to test whether the link is only viable in infancy. Perhaps the limited language knowledge of infants and the strong pressure to acquire new words causes infants to use more of the available information for word learning than adults, who already have large vocabularies and greater cognitive abilities that may obviate these mechanisms. Conversely, it is possible that statistical word segmentation is intrinsically linked to word learning and this link persists into adulthood. Second, testing adults allowed the use of an explicit word learning task rather than inferring word learning indirectly from dis-habituation data. Third, the infant habituation test only showed differences at a single time point, but we can use more flexible tasks with adults to examine possible differences in the trajectory of the learning curve and conduct a finer-grained analysis of the link between statistical segmentation and word learning. Our experiments tested whether adults would be better at learning object labels if the labels were consistent with syllable transition probability (i.e., "words" vs. "partwords"). Experiment 1 tested whether adults learn novel object labels faster when those labels contain high-probability transitions. Experiment 2 further tested whether learning rate differences between "words" and "partwords" are due to inhibition of labels that contain low probability transitions or facilitation of labels that contain high probability transitions.

Experiment 1

Participants

Participants were 49 students at the University of Connecticut who reported English as their native language and normal hearing. They received course credit for participation in the experiment.

Materials

A series of pilot experiments was used to develop a set of auditory materials that replicated the classic statistical segmentation results (Saffran, Aslin, & Newport, 1996; Saffran, Newport, &

Aslin, 1996). The materials were based on syllables spoken by a female native speaker of American English in all possible co-articulatory contexts ("bada", "bapa", "daba", "dapa", etc.). Syllables were recorded in a sound-attenuating room directly to digital format at 22050 Hz. Initial syllables from each two-syllable utterance were then extracted from the recording and matched in duration (301 ms ±1 ms). This duration would produce a speech rate of approximately 200 syllables/minute, but Saffran et al. reported a somewhat faster rate (approximately 216 syllables/minute) and pilot tests suggested that statistical segmentation is more robust at slightly faster speech rates, so the syllables were re-synthesized (without changing pitch) using the PSOLA method in Praat software (Boersma & Weenink, 2005) with a 10% increase in speech rate (corresponding to approximately 219 syllables/minute). Six bisyllabic "words" were created by concatenating syllables using the coarticulatorily appropriate versions of each syllable. Three different syllable streams were created by concatenating four of the words (200 repetitions each, approximately 7 minutes) pseudo-randomly into a running string with no pauses and such that no word occurred twice in a row. In pilot testing ($N = 18$), post-test word identification (see Saffran et al., 1996 for task details) performance following this exposure stream was 61.8% correct (chance = 50%; $t(17) = 3.38$, $p < 0.01$), a level of performance comparable to previous reports (e.g., Saffran et al., 1996).

For the artificial lexicon learning phase, novel black-and-white geometric objects were created by filling eight randomly chosen contiguous cells of a 4×4 grid. Previous studies using these materials and the training paradigm described below found robust word learning, replicated phonological competition results found with real words, and shed light on the development of representations that support spoken word recognition (Magnuson, Tanenhaus, Aslin, & Dahan, 2003).

Procedure

The experiment began with a passive listening exposure phase during which participants listened to a stream of syllables for approximately 7 minutes (as described above). Participants were instructed to listen to a "made-up language" and that they would be asked questions about the language, but were not told that they would be learning words. After the exposure phase, participants completed an artificial lexicon learning phase. Each participant learned four label-object pairings. Participants were randomly assigned to learn labels that contained high probability transitions from the statistical segmentation phase ("words", transitional probability = 100%) or labels that contained low probability transitions from the statistical segmentation phase ("partwords", transitional probability = 33%). Assignment of labels to word/partword condition was counterbalanced across participants.

On each artificial lexicon trial the participants saw two geometric figures (to the left and right of a fixation cross, with target location randomized) and an artificial lexicon item (label) was presented through headphones. Participants responded using the keyboard to indicate whether the object on the left or right corresponded to the spoken item. After the participant's response, visual feedback was presented ("correct" or "incorrect"). The distractor object on each trial was an object corresponding to one of the other items in the artificial lexicon. A training block consisted of presentation of each of the 4 objects with each of the 3 possible distractor objects (12 trials). Participants completed 10 blocks of artificial lexicon learning. The entire experiment was completed in less than 30 minutes.

Results and Discussion

Seven participants (3 from "word" condition, 4 from "partword" condition) were excluded from analyses due to failure to show substantial learning in the artificial lexicon learning phase (overall accuracy less than 55%, chance = 50%). Figure 1 shows the proportion correct responses for the word and partword conditions by block for the remaining 42 participants (21

in each condition). In general, participants learned all object labels relatively well: proportion correct increased gradually from near-chance to near-perfect performance.

There was a clear difference in learning rate. Performance improved faster for the participants learning novel words that contained high probability syllable transitions from the exposure phase than for participants learning novel labels that contained low probability syllable transitions. To quantify this result we used a technique developed specifically for analyzing change over time: growth curve analysis (Singer & Willet, 2003; see also Mirman, Dixon, & Magnuson, in press, for another application of this technique). We used second-order orthogonal power polynomials to capture the curvilinear shape of the learning curves. For these models the intercept reflects average curve height, the linear term reflects the overall slope of the learning curve, the quadratic term reflects the degree of curvature of the learning curve, and condition (word vs. partword) effects on those terms reflect the condition differences in shape of the learning curve. Of particular interest were condition effects on the linear and quadratic terms, which reflect condition differences in learning rate. The base model consisted of the intercept, linear, and quadratic time terms; we then added the condition effects on those terms. Significant improvements in model fit indicate reliable differences between conditions with respect to particular properties of the curve.

There was no significant effect of condition on the intercept ($\Delta LL = 0.1$, *n.s.*), but there were significant effects of condition on the linear ($\Delta LL = 6.0$, $p < 0.01$) and quadratic ($\Delta LL = 5.5$, $p < 0.01$) terms. The full model fitted curves are superimposed on the observed data in Figure 1. These results indicate that there were reliable differences in learning rate between the “word” and “partword” conditions both in terms of overall slope and curvature of the learning curve.

Participants were able to learn the four novel object labels (i.e., meanings) in the span of the experiment: matching a spoken item with the correct novel geometric object started near chance (50%) and increased gradually toward near-perfect performance. This learning progressed faster for participants learning labels that contained high probability syllable transitions (i.e., “words”) from the preceding statistical segmentation phase than for participants learning labels that contained low probability syllable transitions (i.e., “partwords”). One interpretation of these results is that statistical segmentation created memory traces that acted as word candidates, and which were subsequently more readily mapped to meanings. In the partword condition, the artificial lexicon words did not have this statistical segmentation advantage, so these words were learned more slowly. An alternative interpretation is that statistical learning inhibited the mapping of novel labels containing low frequency transitions onto novel objects, thus creating a disadvantage for partwords relative to words. To adjudicate between these alternative interpretations, Experiment 2 added a third, neutral condition to establish a baseline: “nonwords”, which were composed of syllables not presented during the exposure phase.

Experiment 2

Experiment 1 demonstrated that participants learn novel object labels faster when those labels contain high probability syllable transitions than when they contain low probability syllable transitions. In Experiment 2 we examined whether this difference was due to statistical facilitation of high probability syllable transitions or inhibition of low probability syllable transitions. This was done by adding a third condition to the artificial lexicon; in this condition participants learned novel object labels that contained syllables that were not part of the statistical learning exposure stream. Thus, the learning rate for these words provides a baseline that is neutral with respect to statistical segmentation exposure.

Participants

Participants were 93 students at the University of Connecticut who reported English as their native language and normal hearing. They received course credit or a small payment for participation in the experiment.

Design and Procedure

The auditory materials for this experiment were based on the same syllables used in Experiment 1. The 6 possible bi-syllabic "words" were divided into two sets of three "words" and for each set, an exposure stream was created by concatenating 200 repetitions of each word pseudo-randomly, as in Experiment 1. Total exposure time was about 5.5 minutes. For each exposure stream, there were 3 possible artificial lexicon conditions: in the "word" condition participants learned novel object labels that were composed of syllables with high probability transitions in the exposure stream (transitional probability = 100%), in the "nonword" condition the labels were composed of syllables that did not occur in the exposure stream (i.e., they were from the other exposure stream; note that this design makes the assignment of object labels to "word" and "nonword" conditions counterbalanced across participants), and in the "partword" condition the labels were syllables from the exposure stream, but arranged in low transition probability pairs (transitional probability = 50%). Participants were randomly assigned to artificial lexicon condition and exposure stream.

As in Experiment 1, the passive exposure phase was followed by an artificial lexicon learning phase. The artificial lexicon followed the same procedure as in Experiment 1, but because the lexicons were smaller (3 words instead of 4), there were fewer possible unique pairings per block (3 words × 2 possible competitors = 6 trials; in Experiment 1 there were 4 words × 3 possible competitors = 12 trials). Participants completed 10 blocks of artificial lexicon learning; this constitutes fewer learning trials per item than Experiment 1, but since there were fewer items to be learned, the final performance level was approximately equal across the two experiments.

Results and Discussion

Nine participants (4 from "word" condition, 2 from "partword" condition, and 3 from "nonword" condition) were excluded from analyses due to failure to show substantial learning in the artificial lexicon learning phase (overall accuracy less than 55%). Proportion correct responses for word, partword, and nonword conditions for the remaining 84 participants (28 participants per condition) are shown in Figure 2. In all three conditions participants learned the lexicon very well (i.e., there was a gradual increase, reaching near perfect performance at the end of the learning phase). As in Experiment 1, learning was analyzed using growth curve analysis using second-order orthogonal polynomials. There were no significant effects of condition on the intercept ($\Delta LL = 3.1$, $p > 0.1$) or linear ($\Delta LL < 0.1$, *n.s.*) terms, but there was a reliable effect on the quadratic term ($\Delta LL = 6.5$, $p < 0.01$). Because initial and final accuracy were essentially equal across conditions (all participants were guessing at the beginning of learning and mastered the labels by the end of learning), a difference in learning rate was reflected in differences in the curvature of the learning curve. That is, faster learning meant a steeper initial rise and a longer asymptotic plateau, which was captured by differences in the quadratic term in the model. Analysis of parameter estimates for the quadratic term showed that there was no significant difference between the "word" and "nonword" conditions ($t(750) = 0.74$, *n.s.*), but there was a significant difference between "word" and "partword" conditions ($t(750) = 2.1$, $p < 0.05$) and "nonword" and "partword" conditions ($t(750) = 2.7$, $p < 0.01$). There were no other reliable differences between parameter estimates (all $p > 0.1$).

These results replicate the Experiment 1 finding that participants learn object labels faster when they contain high probability transitions than when they contain low probability transitions. In

addition, the finding that labels composed of syllables that were not part of the exposure stream are learned as quickly as the high frequency transition labels suggests that statistical learning inhibits the learning of novel meanings for labels that violate learned transitional probabilities. This could be due to statistically learned "words" interfering with the learning of "partwords" because "partwords" and "words" were composed of the same syllables. On this view "nonwords" were not affected because they were composed of different syllables. In analogous tests, infants were found to learn "word" labels, but not "partword" or "nonword" labels (Graf Estes et al., 2007).

Conclusions

One recent study showed that infants can use statistically segmented words as object labels (Graf Estes et al., 2007). The current study provides several extensions of that finding. First, the present results showed a link between statistical segmentation and word learning in adults who already have large vocabularies; so the linguistic relevance of statistical segmentation is not limited to infants who are just beginning language learning. Second, the present results demonstrated a link between statistical segmentation and word learning using an explicit word learning task, which confirms the conclusion that was previously inferred from dis-habituation results with infants (Graf Estes et al., 2007). Third, the present results suggest a difference in the dynamics of the link between statistical segmentation and word learning for infants compared to adults. For infants, high transitional probability syllable sequences acted as object labels, but low transitional probability and non-occurring sequences did not (there was no dis-habituation effect; Graf Estes et al., 2007). In contrast, adults learned the label-object pairings for both types of utterances, but the high transitional probability and non-occurring utterances were learned more quickly in Experiment 2.

This contrast suggests a developmental difference in the link between statistical segmentation and word learning. For adults, statistical learning appears to inhibit the mapping of labels to novel meanings when the syllable transitions in those labels violate statistically learned transitional probabilities for those syllables. In contrast, Graf Estes et al. (2007) found that for infants statistical learning facilitated the mapping of labels to novel meanings when the syllable transitions in those labels were consistent with statistically learned transitional probabilities. It is likely that this difference between facilitation for infants and inhibition for adults is related to adults' larger vocabularies and phonotactic experience, but the computational basis requires further investigation.

In sum, the results reported here demonstrate a link between statistical segmentation and word learning in adults and suggest a possible developmental difference in this link. Statistical segmentation seems to facilitate learning of favored labels for infants and seems to inhibit learning of disfavored labels for adults.

Acknowledgements

We thank Deirdre Dempsey, Matthew Freiburger, Emma Chepya, and Hillarey Jones for their help with data collection. This research was supported by NIDCD grant R01DC005765 to JSM, NICHD NRSA F32HD052364 to DM and by NICHD grant HD01994 to Haskins Laboratories.

References

Boersma, P.; Weenink, D. Praat: doing phonetic by computer (Version 4.3.22) [Computer program]. 2005. Retrieved September 15, 2005, from from http://www.praat.org/

Graf Estes K, Evans JL, Alibali MW, Saffran JR. Can infants map meaning to newly segmented words? Statistical segmentation and word learning. Psychological Science 2007;18(3):254–260. [PubMed: 17444923]

Magnuson JS, Dixon JA, Tanenhaus MK, Aslin RN. The dynamics of lexical competition during spoken word recognition. Cognitive Science. (in press)

Magnuson JS, Tanenhaus MK, Aslin RN, Dahan D. The time course of spoken word learning and recognition: Studies with artificial lexicons. Journal of Experimental Psychology: General 2003;132 (2):202–227. [PubMed: 12825637]

Mirman D, Dixon JA, Magnuson JS. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. Journal of Memory and Language. (in press)

Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. Science 1996;274(5294): 1926–1928. [PubMed: 8943209]

Saffran JR, Newport EL, Aslin RN. Word segmentation: The role of distributional cues. Journal of Memory and Language 1996;35(4):606–621.

Singer, JD.; Willett, JB. Applied longitudinal analysis: Modeling change and event occurrence. New York: Oxford University Press; 2003.

Stager CL, Werker JF. Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature 1997;388(6640):381–382. [PubMed: 9237755]

Werker JF, Cohen LB, Lloyd VL, Casasola M, Stager CL. Acquisition of word-object associations by 14-month-old infants. Developmental Psychology 1998;34(6):1289–1309. [PubMed: 9823513]

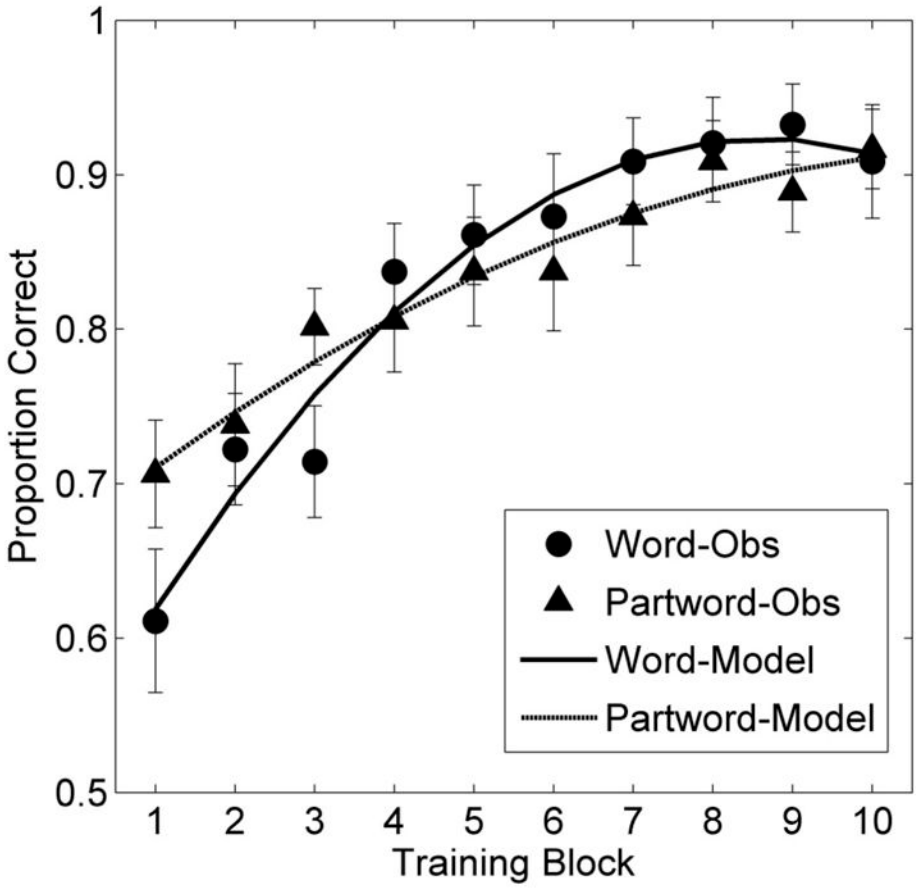

Figure 1.
Experiment 1 proportion correct response by block during the artificial lexicon learning phase. The symbols are observed behavioral data (error bars indicate ± SE), and the lines are growth curve model predictions.

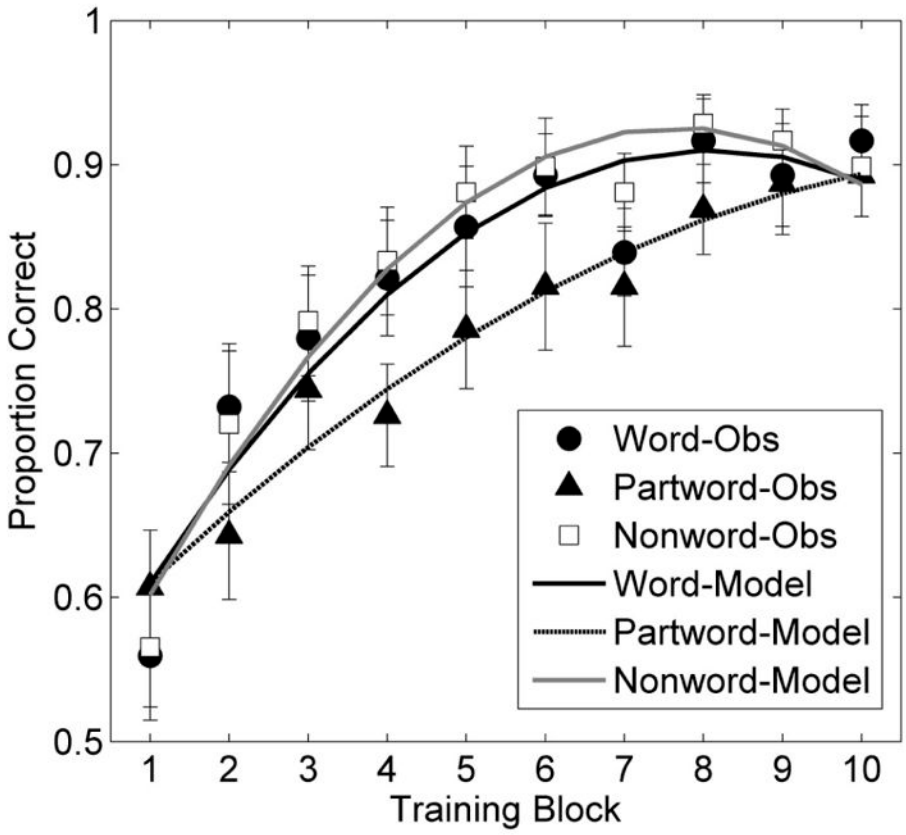

Figure 2.
Experiment 2 proportion correct response by block during the artificial lexicon learning phase. The symbols are observed behavioral data (error bars indicate ± SE), and the lines are growth curve model predictions.