INFANCY, *4*(2), 273–284

BRIEF REPORTS

From Syllables to Syntax: Multilevel Statistical Learning by 12-Month-Old Infants

Jenny R. Saffran and Diana P. Wilson
Department of Psychology & Waisman Center
University of Wisconsin–Madison

To successfully acquire language, infants must be able to track multiple levels of regularities in the input. In many cases, regularities only emerge after some learning has already occurred. For example, the grammatical relationships between words are only evident once the words have been segmented from continuous speech. To ask whether infants can engage in this type of learning process, 12-month-old infants in 2 experiments were familiarized with multiword utterances synthesized as continuous speech. The words in the utterances were ordered based on a simple finite-state grammar. Following exposure, infants were tested on novel grammatical and ungrammatical sentences. The results indicate that the infants were able to perform 2 statistical learning tasks in sequence: first segmenting the words from continuous speech, and subsequently discovering the permissible orderings of the words. Given a single set of input, infants were able to acquire multiple levels of structure, suggesting that multiple levels of representation (initially syllable-level combinations, subsequently word-level combinations) can emerge during the course of learning.

Months before they speak their first words, infants are engaged in rapidly acquiring their native language. To uncover the mechanisms by which this learning occurs, researchers create novel linguistic systems designed to tap particular learning capacities; these materials often take the form of artificial languages that mirror some aspects of natural languages (for review, see Gomez & Gerken,

Supplementary materials to this article are available on the World Wide Web at http://www.infancyarchives.com.

Requests for reprints should be sent to Jenny R. Saffran, Department of Psychology, University of Wisconsin–Madison, Madison, WI 53706. E-mail: jsaffran@wisc.edu

2000). Young infants adeptly learn novel linguistic materials created and presented in the laboratory, from phonemic contrasts (Maye, Werker, & Gerken, 2002) to phonological patterns (Saffran & Thiessen, 2003) to the discovery of word boundaries (e.g., Echols, Crowhurst, & Childers, 1997; Goodsitt, Morgan, & Kuhl, 1993; Johnson & Jusczyk, 2001; Saffran, Aslin, & Newport, 1996) to rudimentary syntax (e.g., Gomez & Gerken, 1999; Marcus, Vijayan, Bandi Rao, & Vishton, 1999).

Each of these experiments was designed to isolate infant learning mechanisms by creating learning experiences and test stimuli that target a specific level of analysis (e.g., either word segmentation or syntax). Given real language input, however, the child could, in principle, apply any number of learning mechanisms to the input corpus. What tells the learner which mechanism is actually the "right" one for a particular learning problem, given multiple possible mechanisms? Does the learner know that certain types of input should be processed by particular mechanisms? Or does the input drive the learning process, such that all potentially pertinent mechanisms are applied to the input, with a winner emerging after computations have been performed? Or does one mechanism feed the next, with a complex interaction between the learner's inherent biases and the structure of the input? In the current research, we address the latter hypothesis—the capacity to apply learning mechanisms sequentially such that the output of one serves as the input to the next—by using exposure and test materials designed to assess the acquisition of multiple aspects of language during a single experimental session.

Input to language learners does not contain transparent labels marking eventual levels of structure. Consider the sound string *datopidu*. Is this sequence a single word, or does it contain multiple words? If the latter, are the words organized according to a syntax? Statistical cues derived from the distribution of these sounds across multiple utterances would be useful to the learner in answering these questions, because infants can track sequences of sounds to discover word boundaries (e.g., Saffran et al., 1996) and sequences of words to discover rudimentary syntax (e.g., Gomez & Gerken, 1999). Although language contains multiple levels of structure, the input is not organized to point learners to different levels at different times; it is not the case that infants initially receive input prespecified for the discovery of word boundaries and later receive a different set of input prespecified for syntax acquisition. Instead, learners must be able to use output from one learning mechanism as input to additional learning mechanisms to discover multiple levels of structure within a single set of input materials. This has not been reflected in prior experiments, wherein, for example, infants engaged in syntax learning tasks receive input containing pauses between words, which removes word segmentation as a component of the task (e.g., Gomez & Gerken, 1999; Marcus et al., 1999).

In previous research, we have asked whether the output of statistical language learning takes the same representational format as the input (Saffran, 2001). When learners are first confronted by a sequence like *datopidu* as part of an extended stream of fluent speech, they presumably initially represent sounds, such as syllables, as primitives, which then enter into the language learner's statistical computations. Following learning, however, the primitives are no longer the individual sounds. Instead, the primitives become the words—*dato* and *pidu*—that are stored and available to enter into subsequent computations as part of the native language, as reflected by infants' behavior when these words are later embedded in familiar English sentence frames (Saffran, 2001).

Once lexical representations emerge as a function of word segmentation, an array of new possible computations awaits the language learner. In particular, the learner is now in a position to detect patterns of words. Prior to word segmentation, multiword patterns are presumably opaque. To discover rudimentary syntax, learners must first find words; the syntactic patterns that generated the sequence *datopidu* are invisible until the learner discovers that *dato* and *pidu* are distinct units. For this process to work correctly, the infant must perform two different sets of computations over the same input—first finding the patterns of sublexical units that cohere into words, and then finding the regularities governing the lexical units. The output of the first process thus serves as input to the second.

Experiments 1 and 2 test the hypothesis that the statistical learning process generates lexical units from sublexical units, which can then enter into the detection of syntactic regularities. If infants can perform both types of computations across the same set of input (detecting patterns of sublexical units for word segmentation, and subsequently detecting patterns of lexical units for syntax learning), this would suggest that infant learning mechanisms can operate over representations derived not only from the environment, but also from internal learning processes. The words that are the product of the first learning mechanism, and that serve as input to the second, certainly reflect systematic patterns in the input, but they are not actually present in the input; instead, they are representations manufactured by the infant's learning mechanisms.

We thus combined a word segmentation paradigm (e.g., Saffran et al., 1996) with a finite state grammar learning task (e.g., Gomez & Gerken, 1999). Twelve-month-old infants (an age at which infants continue to be engaged in word segmentation, and are beginning to acquire rudimentary syntax) were exposed to multiword sentences generated by a simple grammar. Importantly, there were no acoustic cues to the boundaries between the multisyllabic words within the sentences; only statistical cues to word boundaries were present. For example, the sentence *dato pidu buto badu dipa* was presented as *datopidubutobadudipa*. Thus, to learn the grammar, the infants must first segment the words. Infants were then tested on novel grammatical and ungrammatical sentences. Critically, the test was constructed such that infants could not discriminate between the two types of test items based on syllable-level information; the transitional probabilities between syllables were identical in the grammatical and ungrammatical sentences. Instead, test discrimination required higher order knowledge of permissible word sequences in the grammar. Success would indicate that infants used the output of one learning mechanism as the input to another learning mechanism, with no changes in the external stimuli driving the learning process.

EXPERIMENT 1

Infants were exposed to a list of sentences, each consisting of five disyllabic words with an optional sentence-initial syllable. The materials were synthesized as fluent speech, with a brief pause after each sentence. Within each sentence, the words were organized according to a finite-state grammar. Following exposure, infants were tested on novel sentences that were either grammatical or ungrammatical. The languages were constructed such that both types of test items contained equal transitional probabilities between syllables (all 1.0 within words and .25 at word boundaries). Thus, test discrimination could not be based on syllable-pair transitional probabilities. Similarly, absolute syllable position was not a cue for discrimination due to the optional initial syllable, which ensured that infants could not succeed by noticing the absolute positions of particular syllables. Instead, discrimination required knowledge of permissible word orderings.

Method

Participants. Forty full-term 11- to 12-month-old monolingual infants with no history of recurrent ear infections were tested (mean age = 11 months, 27 days; range = 11:1 to 12:14). Infants were randomly assigned to one of two counterbalanced exposure languages (A and B). Fifty-six additional infants were tested but not included in the analysis for the following reasons: fussiness (22), parental interference (13), fewer than eight total test trials completed when enforcing a minimum listening time per trial of 2.5 sec (9), looking times averaging less than 3 sec to one or both sides (6), not looking at the side lights (3), exceeding the maximum time of 10 min allowed for the test (2), and experimental error (1).[1] All infants in this and the subsequent experiment were solicited from local birth announcements, and parental consent was obtained prior to testing in accordance with the guidelines of the local human subjects review committee and the principles of ethical treatment established by the American Psychological Association.

[1]We suspect that the high dropout rate reflects disinterest in the synthesized speech on the part of 12-month-old infants. This is supported by the large number of infants whose listening times were too brief to be included. Other studies using this age group (e.g., Gomez & Gerken, 1999) included a 5-min break between familiarization and testing, which may assist in participant retention with this age group.

Stimuli. For the familiarization phase, each language consisted of 16 sentences repeated twice. Each sentence was comprised of five disyllabic nonsense words (see Appendix A). Within each language, all sentences conformed to a finite-state grammar. As shown in Table 1, the grammars for both languages were identical other than the positions of the B and D words, which are opposite. Additionally, half of the sentences in each language were preceded with the syllable /la/ to prevent infants from learning the absolute positions of syllables in the sentences.

Each of the four test items consisted of a set of five novel sentences that were either grammatical or ungrammatical relative to the familiarization materials (see Appendix B). The ungrammatical test items for Language A were the grammatical test items for Language B, and vice versa. This between-subject counterbalanced design ensured that any observed preference for grammatical or ungrammatical sentences across the two languages was due to learning.

The materials were constructed such that the transitional probabilities between syllables were identical for the grammatical and ungrammatical sentences. This was accomplished by using the same final syllables for the A and C words and for the B and D words. For example, the test string *dato pidu buto badu dipa*, which was grammatical for Language A but ungrammatical for Language B, consisted of the same syllable-pair transitional probabilities relative to both input grammars (1.0 within words, .25 at word boundaries). Thus, transitional probabilities between syllables were not a cue to grammatical structure.

The languages and test items were generated by a speech synthesizer (MacinTalk) in a monotone female voice. There were no acoustic or prosodic markers at word boundaries, other than a 1.4-sec pause between sentences. The speech was digitized for playback during the experimental session. Languages A and B each lasted 2 min 8 sec and each test item lasted 18 sec. The audio files for Languages A and B, and for the four test items, are available at http://www.infancyarchives.com.

Procedure. Infants were tested individually while seated in a parent's lap in a sound-attenuated booth. An observer outside the booth monitored the infant's

TABLE 1
Finite State Grammars and Vocabularies for Languages A and B

Language A grammar: A → B → C → D → E
Language B grammar: A → D → C → B → E
Vocabulary:
A: *dato, kuga*
B: *pidu, gobi*
C: *buto, tiga*
D: *badu, tubi*
E: *dipa, tako*

looking behavior on a closed-circuit TV system and coded the infant's behavior using a button-box connected to a PC. This button-box was used to initiate trials and to enter the direction of the infant's head turns, which controlled the duration of each test trial. Both the parent and the observer listened to masking music over headphones to eliminate bias.

At the beginning of the familiarization phase, the infant's gaze was first directed to a blinking light on the front wall. Then the sound sequence for one of the two languages was presented from two loudspeakers (one located on each of the two side walls). During this familiarization period, to keep the infants' interest, a blinking light above one of the two loudspeakers (randomly selected) was lit and extinguished dependent on the infant's looking behavior. When this side light was extinguished, the central blinking light was illuminated until the infant's gaze returned to center, and another blinking side light was presented to elicit the infant's gaze. During the familiarization phase there was no contingency between lights and sound, which played continuously.

Immediately after familiarization, 12 test trials were presented (3 trials for each of the 4 test items, presented in random order). All infants heard the same 12 trials: 6 grammatical sentence sets and 6 ungrammatical sentence sets. Each test trial began with the blinking light on the front wall. When the observer signaled the computer that the infant was fixating this central light, one of the lights on the two side walls began to blink and the central light was extinguished. When the observer judged that the infant had made a head turn of at least 30° in the direction of the blinking side light, a button press signaled to the computer that one of the test items should be presented from the loudspeaker adjacent to the blinking light. This test item played until the observer coded the infant's head turn as deviating away from the blinking light for 2 consecutive sec. When this look-away criterion was met, the computer extinguished the blinking side light, turned off the test stimulus, and turned on the central blinking light to begin another test trial. The computer randomized the order of test trials (three for each of the four test items) and accumulated total looking time to each test item. Test trials automatically ended after a single repetition of the sentence set.

RESULTS AND DISCUSSION

Prior to analyzing the data, we excluded trials with listening times of less than a 2.5-sec duration, because the head turn on these trials was initiated prior to the completion of the first sentence of the set. This criterion led to the exclusion of 59 of the 480 test trials (12%). A two-factor analysis of variance (ANOVA) assessed whether infants exposed to the two familiarization languages discriminated the novel grammatical sentences from the novel ungrammatical sentences. Although there was no difference in listening times between the two counterbalanced languages, $F(1, 38) = .15$, *ns*, there was a significant difference in listening times for grammatical versus ungrammatical sentences, $F(1, 38) = 5.37$, $p < .05$; the interaction was not significant, $F(1, 38) = .36$, *ns*. As shown in Figure 1, infants listened longer to grammatical sentences ($M = 7.34$ sec, $SE = .34$ sec) than to ungrammatical sentences ($M = 6.55$ sec, $SE = .34$ sec). These results suggest that the infants learned the simple grammatical patterns that characterized the exposure materials. To do so, infants must have first used syllable-level statistics to isolate words, and then determined which words followed which other words. Had the infants represented the test materials solely in terms of transitional probabilities between syllables, no discrimination could have occurred, because the grammatical and ungrammatical sequences contained identical syllable-pair transitional probabilities.

Although these results suggest that infants can track multiple levels of regularities over the same set of input, the effect was relatively small—only 24 of the 40 infants preferred the grammatical sentences—and a large number of infants were excluded from the analyses. We thus ran a second experiment designed as a replication of Experiment 1. In an attempt to simplify the task for the infants, we removed the sentence-final word from the grammar, which served to decrease the number of words that infants were required to learn. Otherwise, the methods were identical to Experiment 1.

EXPERIMENT 2

Method

Participants. Forty 11- to 12-month-old monolingual infants were tested (mean age = 11 months, 17 days; range = 11:3 to 12:12). Forty-four additional infants were tested but not included in the analysis for the following reasons: fussiness

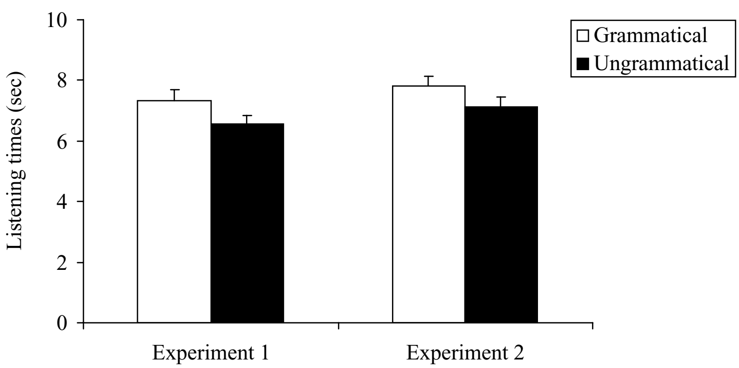

FIGURE 1 Listening times for grammatical and ungrammatical sentences in Experiments 1 and 2.

(15), looking times averaging less than 3 sec to one or both sides (12), fewer than eight total test trials completed when enforcing a minimum listening time per trial of 2.5 sec (7), parental interference (7), exceeding the maximum time of 10 min allowed for the test (2), and experimental error (1).

Stimuli. The languages and test items used in Experiment 1 were edited to remove the sentence-final word. Languages A and B each lasted 1 min 48 sec, and each test item (of the four sets of test sentences) lasted 15 sec.

Procedure. The procedure was identical to Experiment 1.

Results

As in Experiment 1, trials with listening times of less than a 2.5-sec duration were excluded from the analyses, removing 54 of the 480 test trials (11%). A two-factor ANOVA assessed whether infants exposed to the two familiarization languages discriminated the novel grammatical sentences from the novel ungrammatical sentences. Although there was no difference in listening times between the two counterbalanced languages, $F(1, 38) = .02$, *ns*, there was a significant difference in listening times for grammatical versus ungrammatical sentences, $F(1, 38) = 4.67$, $p < .05$; the interaction was not significant, $F(1, 38) = .02$, *ns*. As shown in Figure 1, infants listened longer to grammatical sentences ($M = 7.80$ sec, $SE = .34$ sec) than to ungrammatical sentences ($M = 7.14$ sec, $SE = .28$ sec); this pattern was shown by 25 of the 40 infants. Removing the final word from the input sentences thus did not simplify the learners' task; these results do not differ from those of Experiment 1. The results of the two experiments thus converge to support the hypothesis that infants can engage in two learning processes, segmentation and syntax learning, with the output of the first process serving as input to the second.

GENERAL DISCUSSION

To develop procedures designed to uncover infant learning mechanisms, researchers generally study only a single aspect of language learning in any given experiment. Conflicting cue designs are often used to ask how infants prioritize different types of information, but again, these studies typically involve one linguistic task, such as word segmentation (e.g., Johnson & Jusczyk, 2001; Mattys, Jusczyk, Luce, & Morgan, 1999; Morgan, 1996; Morgan & Saffran, 1995; Thiessen & Saffran, 2001). The relationship between learning mechanisms operating at different levels of linguistic structure has thus been difficult to explore.

In the current experiments, infants were confronted by two linguistic tasks: segmenting words from continuous speech and discovering the ordering relationships between those words. Prior to accomplishing word segmentation, the grammatical structure of the language was not accessible to the learners. Infants were then tested on new sentences that were either grammatical or ungrammatical with respect to the input grammar. Infants could not have performed the test discrimination based on syllable-level information, because the transitional probabilities between syllables were identical for the grammatical and ungrammatical sentences. Instead, the infants' successful performance indicated that they moved beyond the level of syllable-pair probabilities to track the relationships between the words themselves.

We thus suggest that learners performed two different sets of statistics on the input, first tracking the transitional probabilities between syllable tokens to find words, and then tracking the distributions of the words, such that sentences not previously encountered were discriminated based on their consistency with the input grammar. One alternative hypothesis is that the infants engaged in second-order probability learning. That is, rather than tracking the probabilities of syllable pairs, infants tracked the probabilities of syllable triples. Consider the sequence *datopidu*, which was grammatical for infants in Condition A but ungrammatical for infants in Condition B. In terms of syllable-pair probabilities, the two conditions are identical; in both cases, the sequence of probabilities is 1.0 for *dato*, .25 for *topi*, and 1.0 for *pidu*. However, from the perspective of syllable triples, the probability that *dato* is followed by *pi* is .5 for Condition A, but 0 for Condition B. Thus, infants could have successfully discriminated the two types of test sentences based on second-order probabilities. However, we consider this possibility unlikely because in prior experiments, infants exposed to syllable-pair probabilities like 1.0 followed by .25 used the trough in the probability function as a segmentation cue (e.g., Aslin, Saffran, & Newport, 1998; Saffran, 2001; Saffran et al., 1996). That is, infants do not treat "part-words" that span word boundaries (as defined statistically) in the same way that they treat words. To use the second-level statistical information in the input, infants would have had to ignore the syllable-pair statistical cues to word boundaries, which is unlikely given the results from prior experiments. We thus suggest that infants did indeed perform the test discrimination by attending to the word sequences rather than to *n*-order syllable-level probabilities.

A second alternative hypothesis concerns the use of longer sequences of syllables to discriminate grammatical and ungrammatical sentences. When one considers sequences of four or six syllables, it is necessarily the case that the grammatical sentences are more similar to the familiarization sentences than the ungrammatical sentences. Could infants have responded to the test sentences on this basis; for example, noticing that they had heard *butobadudipa*, but not *butogobidipa*, during exposure? Although this is possible, it is unlikely, given that as discussed previously, infants use the probabilities with which syllables co-occur to segment continuous speech into shorter sequences that correspond to the statistics of the syllable combinations (e.g., Aslin et al., 1998; Saffran, 2001; Saffran et al., 1996). This is the case even when syllable pairs are matched for frequency, that is, when words and sequences that span word boundaries occur equally often in the input, but where the words contain higher internal probabilities than the sequences spanning word boundaries (Aslin et al., 1998). As argued earlier, it seems unlikely that these learners would have ignored the strong segmentation cues in the input, given infants' previously reported capacity to automatically and rapidly use this type of information in segmentation tasks, and instead encoded longer sequences unsupported by the statistics of the input.

Infants began this task without any knowledge of the syllable combinations relevant to this learning task; their initial representations on hearing these stimuli were presumably the sounds themselves. Following learning, infants ended the task with word-level representations organized by a simple grammar. Such representational change is necessary for natural language acquisition, during which learners must draw generalizations across multiple levels of language. These studies suggest that the output of one learning process can serve as internally generated input for a second learning process, representing an early step in understanding how infants' arsenal of learning mechanisms interact with one another. Despite the relative simplicity of the materials used in this task, the fact that infants moved from syllables to rudimentary syntax following a few minutes of exposure suggests that the learning processes under investigation are not only powerful, but are also well-designed for solving the problems facing human language learners.

ACKNOWLEDGMENTS

This research was funded by grants from NICHD (RO1HD37466) and NSF (BCS-9983630). We thank Maria Boardman, Susan Bunton, Michelle Loman, Joanna Miller, Aimee Niehbur, Kari Reeck, AnnMarie Rian, Rachel Robertson, Erik Thiessen, and Cori Yahnke for their assistance in the conduct of this research, and to three anonymous reviewers for helpful comments on a previous version of the article.

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.

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). Perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language, 36,* 202–225.

Gomez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition, 70,* 109–135.

Gomez , R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4,* 178–186.
Goodsitt, J. V., Morgan, J. L., & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. *Journal of Child Language, 20,* 229–252.
Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44,* 548–567.
Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283,* 77–80.
Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38,* 465–494.
Maye, J., Werker, J. F., & Gerken, L. (2002). *Infant sensitivity to distributional information can affect phonetic discrimination. Cognition, 82*, B101–B111.
Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language, 35,* 666–688.
Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and supersegmental information in preverbal speech segmentation. *Child Development, 66,* 911–936.
Saffran, J. R. (2001). Words in a sea of sounds: The output of statistical learning. *Cognition, 81,* 149–169.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.
Saffran, J. R., & Thiessen, E. D. (2003). *Pattern induction by infant language learners.* Manuscript under review.
Thiessen, E., & Saffran, J. R. (2001, April). *Infant use of stress cues versus statistical cues to word segmentation at 7, 9, and 11 months.* Poster presented at the biennial meeting of the Society for Research in Child Development, Minneapolis, MN.

APPENDIX A

Language A	*Language B*
dato pidu buto badu dipa.	la kuga tubi tiga pidu tako.
kuga gobi buto tubi tako.	dato badu buto pidu dipa.
la dato gobi buto badu tako.	la dato badu buto gobi tako.
kuga gobi tiga tubi dipa.	kuga tubi tiga gobi dipa.
la dato pidu buto tubi tako.	kuga badu buto gobi dipa.
dato gobi buto tubi dipa.	dato tubi tiga gobi tako.
la kuga gobi tiga badu tako.	la kuga tubi buto pidu dipa.
dato gobi tiga tubi tako.	la dato badu tiga gobi dipa.
la kuga gobi buto badu dipa.	kuga badu tiga pidu dipa.
kuga pidu buto tubi dipa.	la kuga badu buto pidu tako.
la dato pidu tiga badu tako.	dato tubi buto gobi dipa.
kuga pidu tiga tubi tako.	la dato badu tiga pidu tako.
la dato gobi tiga badu dipa.	dato tubi tiga pidu dipa.
la kuga pidu buto badu tako.	kuga tubi buto gobi tako.
dato pidu tiga tubi dipa.	la kuga badu tiga gobi tako.
la kuga pidu tiga badu dipa.	la dato tubi buto pidu tako.

APPENDIX B

Test Items	*Language A*	*Language B*
kuga pidu buto badu dipa	Grammatical	Ungrammatical
dato gobi tiga badu tako		
kuga gobi tiga tubi tako		
dato pidu buto tubi dipa		
dato pidu buto badu tako		
dato gobi buto tubi tako	Grammatical	Ungrammatical
kuga gobi buto tubi dipa		
kuga pidu tiga badu tako		
dato pidu tiga badu dipa		
kuga gobi tiga tubi tako		
kuga tubi buto gobi dipa	Ungrammatical	Grammatical
dato badu tiga gobi tako		
dato tubi buto pidu dipa		
kuga badu tiga pidu tako.		
dato badu buto pidu tako		
dato badu tiga pidu dipa	Ungrammatical	Grammatical
kuga tubi tiga gobi tako		
kuga badu buto pidu dipa		
dato tubi buto gobi tako		
kuga tubi tiga gobi tako		