ARTICLE IN PRESS

Cognition xxx (2008) xxx–xxx

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Brief article

Learn locally, act globally: Learning language from variation set cues

Luca Onnis [a,b,*], Heidi R. Waterfall [c,d], Shimon Edelman [c]

[a] Department of Second Language Studies, University of Hawaii, 1890 East-West Road, Honolulu, HI 96822, United States
[b] Center for Second Language Research, University of Hawaii, 1890 East-West Road, Honolulu, HI 96822, United States
[c] Department of Psychology, Cornell University, Uris Hall, Ithaca, NY 14850, United States
[d] Department of Psychology, 5848 S. University Avenue, Chicago, IL 60637, United States

ARTICLE INFO

Article history:
Received 4 June 2008
Revised 11 September 2008
Accepted 3 October 2008
Available online xxxx

Keywords:
Variation sets
Artificial language learning
L1
L2
Language acquisition
Artificial grammar learning
Corpus analyses
Alignment
Short-term memory

ABSTRACT

Variation set structure – partial overlap of successive utterances in child-directed speech – has been shown to correlate with progress in children's acquisition of syntax. We demonstrate the benefits of variation set structure directly: in miniature artificial languages, arranging a certain proportion of utterances in a training corpus in variation sets facilitated word and phrase constituent learning in adults. Our findings have implications for understanding the mechanisms of L1 acquisition by children, and for the development of more efficient algorithms for automatic language acquisition, as well as better methods for L2 instruction.

1. Variation sets in language learning

Imagine receiving two messages from outer space (Table 1). While message 1 appears hard to decipher, message 2 reveals certain structural properties of the unfamiliar language. By aligning and comparing its successive utterances (1b–2b, 2b–3b, and 3b–4b), one notices that some but not other sequences repeat, suggesting potential segmentation into words (Table 2, left). What promotes language discovery in (1b–4b), but not in (1a–4a), is that some of the successive utterances in that passage form partial self-repetitions, or *variation sets* (Küntay & Slobin, 1996).

Harris (1946) proposed *alignment* and *comparison* of successive utterances as simple procedures for linguistic inquiry. There is reason to believe that these processes may also facilitate language learning in children (L1) and adults (L2). Variation set structure is prevalent both in child-directed speech and in adult conversations (e.g., Hoff-Ginsberg, 1986; Pickering & Garrod, 2004; Szmrecsanyi, 2005), and it correlates with progress in children's acquisition of syntax (e.g., Hoff-Ginsberg, 1986; Waterfall, 2006, submitted for publication-b). Indeed, our second example (1b–4b) is patterned on a passage from a real corpus of child-directed speech (Waterfall, submitted for publication-a; Table 2, right).

From a computational standpoint, the key characteristic of variation sets is that *local* mechanisms of alignment and comparison allow even memory-limited learners to discover structure that they would otherwise miss. The same mechanisms can be reused to yield higher-order structural properties of language (Table 2, right). But just how frequent are variation sets in natural child-directed speech (CDS)? Researchers have reported that about 20% of utterances in the corpora they studied appeared within

* Corresponding author. Address: Department of Second Language Studies, University of Hawaii, 1890 East-West Road, Honolulu, HI 96822, United States.
E-mail address: lucao@hawaii.edu (L. Onnis).

0010-0277/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2008.10.004

Table 1
Two hypothetical language learning situations.

Alien message #1	Alien message #2
(1a) kedmalburafuloropesai	(1b) kedmalburafuloropesai
(2a) gianaber	(2b) rafuloro
(3a) manadukbiunel	(3b) manaloropesai
(4a) kiciorudanamjeisulcaz	(4b) kedmalbumanaloropesai

Table 2
Alignment of consecutive utterances can reveal a great deal of structure. The overlapping material across the first two sentences, *push them*, is a verb phrase. Alignment and comparison would break the first sentence into constituents such as a verb phrase (*push them*), and a prepositional phrase (*to school*).

Alien message #2 aligned	Sample of English child-directed speech
(1c) kedmalbu rafu loro pesai	You got to push them to school
(2c) ------ rafu loro ---	------ Push them -----
(3c) ------ rafu loro pesai	------ Push them to school
(4c) ------ mana loro pesai	------ Take them to school

variation sets (e.g., Küntay & Slobin 1996; Newport, Gleitman, & Gleitman, 1977; Waterfall, 2006). To corroborate this figure, we analyzed 70,208-utterances of CDS in the Lara corpus, one of the densest longitudinal collections of natural conversation transcripts of a mother and child between 1;9 and 3;3 years of age (Rowland, Pine, Lieven, & Theakston, 2005). We found that 58.6% of successive sentences shared at least one word.[1] Moreover, 34.9% of unique words in the corpus appeared at least once in the matching portion of some variation set.[2] Monte Carlo simulations showed these figures to be significantly higher than chance (cf. Fig. 1).

Given the apparent prominence of variation sets in child-directed speech, we set out to test their effect on artificial language learning by adults, investigating whether arranging 20% of utterances in variation sets would facilitate segmenting words in continuous speech (Experiment 1) and identifying phrasal constituents (Experiment 2).[3]

2. Experiment 1: learning word segmentation

Experiment 1 tested the subjects' ability to segment continuous speech into word-like units.

2.1. Participants and materials

Fifty-four Cornell students were paid $4. In the learning phase, participants listened to a sequence of utterances ("sentences") consisting of concatenated pseudowords, generated by a rewrite rule: S → A B C. The classes A, B, and C can be conceived of as lexical categories containing respectively three, four, and three lexical items (Table 3). The language generated a total of $3 \times 4 \times 3 = 36$ unique sentences.

For sentence generation, spaces were removed (e.g., S → da pera guklozi → daperaguklozi), and letters were mapped onto phonemes. Synthetic text-to-speech conversion with MBROLA (Dutoit, 1997) generated a seamless stream of phonemes (80 ms for consonants and 260 ms for vowels) for each sentence. Only sentence boundaries (800 ms pauses) signalled word boundaries for the A and C words at the edge of sentences. The Italian diphone set in MBROLA gave participants the impression of being engaged in a novel language learning task. All phonemes had an equivalent phonemic realization in English and were thus familiar enough to English speakers (no participant had studied a Romance language).

A total of 106 sentences were presented via headphones during the learning phase. Because the A and C words varied in syllable length (1–3 syllables), variability in sentence length (4–10 syllables) ensured that participants did not base segmentation on word length.

There were two learning conditions, Varset and Scrambled, which consisted of *exactly the same 106 sentences* but crucially differed in their order of presentation. The Scrambled condition served as control: sentences were pseudo-randomly ordered such that no two adjacent sentences shared any lexical items (Table 3). This condition established a baseline for learning in the absence of variation set structure.

In contrast, sentences in the Varset condition were pseudo-randomly ordered such that 20% of adjacent sentences contained one overlapping lexical item.[4] The remaining 80% satisfied the same criterion as in the Scrambled condition (no lexical overlap between adjacent sentences). Varset and Scrambled sentences alternated in blocks: 6 blocks of 4 Varset sentences and 6 blocks of 13 or 14 Scrambled sentences (Table 3).

Testing involved a forced choice between words and part-words, the latter defined as syllable sequences straddling word boundaries. Participants who succeeded in segmenting the words in the learning phase should reliably prefer words over part-words. Test words were the four words that never appeared at the edge of sentences, and thus were 'buried' in the speech stream without acoustic cues signalling their boundaries (B words: {pera, kadro, fama, zupa}). As test part-words, we chose 8 out of the 20 bisyllabic segments that straddled word boundaries. For instance, in the segment kosipera, sipe was a part-word formed by the last syllable of ko**si** and the first syllable of **pe**ra. Acoustically, words and part-words were equally good word candidates, because (i) during the learning phase they were part of a seamless chain of phonemes, and (ii) they were resynthesized as units for the test.

Words and part-words had different statistical properties. Words had internal transitional probabilities between

[1] Very high frequency closed class words as well as interjections were excluded: a, the, 's, 't, 're, 've and um.

[2] Waterfall (2006) used a more stringent criterion for overlap, which allowed only open-class words to anchor a variation set. Indeed, using the same criterion, 27.9% of the sentences in the Lara corpus were found to be in variation sets.

[3] The choice of 20% was meant to reproduce the most conservative proportion found in corpora so far.

[4] We selected this more conservative percentage based on previous studies (e.g., Küntay & Slobin 1996; Waterfall, 2006) and on our own analyses of the Lara corpus considering only open-class words (nouns, verbs, adjectives, and adverbs) as overlapping between adjacent sentences.

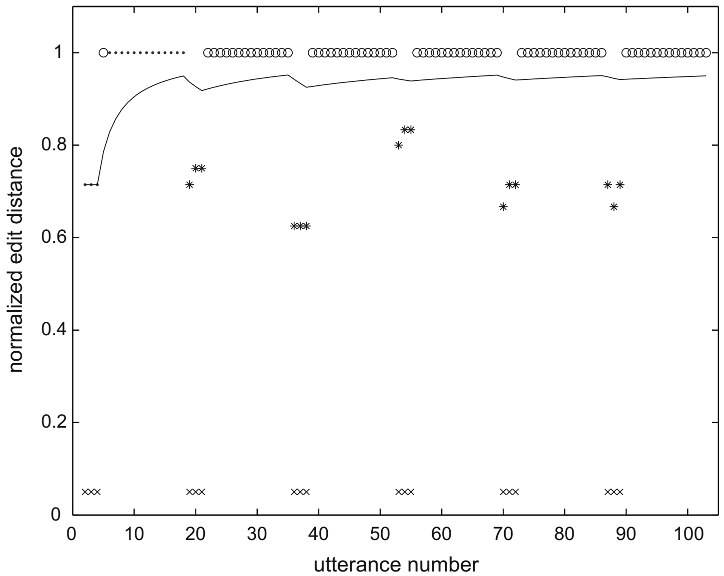

Fig. 1. A variation set can be characterized by its length in utterances (2.88 ± 0.012 mean and standard error in the Lara corpus), and also by the dissimilarity of its members. Dissimilarity can be measured by the Levenshtein (edit) distance; given two sequences of words, it is defined as the smallest number of (possibly individually weighted) elementary edit operations – insertions, deletions, and substitutions of words – that transform one into the other. Dividing this number by the length of the longer sequence normalizes edit distance to lie between 0 and 1. In the Lara corpus, the mean variation set diameter in terms of the normalized edit distance was .82 (se = .002). To assure safe generalization, any corpus-based inference about structure entertained by the learning algorithm needs to pass a test of statistical significance (Edelman & Waterfall, 2007). Given a variation set, the null hypothesis is that of a chance partial alignment of utterances. The learner may test it by comparing the dissimilarity between the utterances to a baseline value – e.g., the cumulative average dissimilarity for the corpus at hand. An analysis of edit distances between successive sentences in the training data in the Varset condition reveals that in almost every variation set, the edit distance between the two sentences is significantly smaller than the baseline provided by the cumulative average. Edit distances d_n between successive utterances (n and n+1) in the Varset training data of Experiment 1 are plotted against n. Solid line: cumulative average $d_{avg} = (1/n)\sum_{i=1}^{n} d_i$. (·): pairs for which d_n and d_{avg} do not differ significantly according to a 2-sided t-test. (*): $d_n < d_{avg}$. (o): $d_n > d_{avg}$. (×) at the bottom of the plot denotes alignable pair. Note that the cumulative statistics of the edit distance values reveal most of the alignments, where they exist, to be significant. A learner can rely on this feature of the training corpus in distinguishing between significant and spurious patterns in structure discovery.

Table 3
Materials and sample sentences presented during the learning phase in Experiment 1.

Structure and lexicon	Scrambled Condition	Varset Condition
S → A B C	kosifamapiu	daperaguklozi
A = {da, kosi, spinose}	spinoseperaguklozi	kosiperapiu
B = {pera, kadro, fama, zupa}	kosikadropiu	spinoseperaguklozi
C = {piu, prati, guklozi}	daperaguklozi	daperaprati

syllables of 1, while part-words had mean internal transitional probabilities of 0.30 (e.g., $TP(pe|si)$ = 0.25). In addition, the test words had a much higher frequency (mean = 25.75, sd = 0.96) than the part-words (mean = 8.5, sd = 0.51). As both adults and infants are sensitive to loci of high and low transitional probability and use them to group syllables into word-like units (Saffran, Aslin, & Newport, 1996), we anticipated that the words in this language could be discovered and preferred over part-words in either condition (see also Fig. 1).

The four test words were presented twice in counterbalanced order, once in pair with a part-word that contained the first syllable of the test word, and once with a part-word that contained its second syllable. For each pair, participants had to chose which one was a word. Two of the test words constituted overlapping items in variation sets during training (In words), and two did not (Out words) in counterbalanced order across subjects (InOut factor). This manipulation allowed us to determine whether the effect of variation sets extended to the segmentation of words that were not directly experienced in variation sets.

2.2. Procedure

Participants were asked to listen to a miniature language containing new words in running speech, a situation akin to a child learning its first language, or to an adult learning a foreign language. They were told they would be tested later. They were randomly and blindly assigned to either the Varset or Scrambled condition. Learning lasted 5 min. In each test trial, participants had to choose the word over the part-word out of the two presented sound sequences.

2.3. Results

Participants in the Varset condition preferred words over part-words on average 5.57 times out of 8, which is

Please cite this article in press as: Onnis, L., et al. Learn locally, act globally: Learning language from variation set cues. *Cognition* (2008), doi:10.1016/j.cognition.2008.10.004

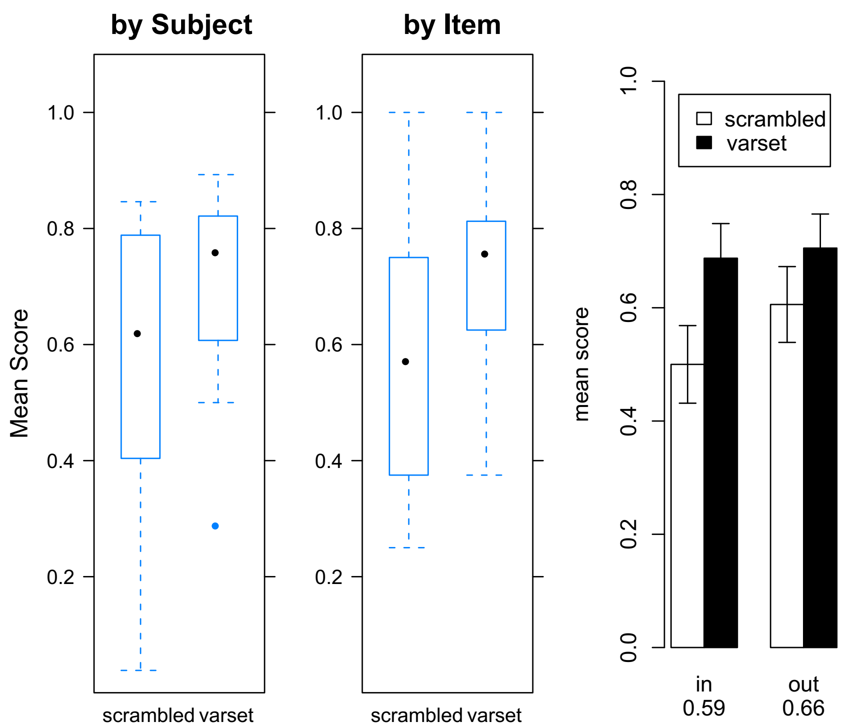

Fig. 2. Experiment 1. *Left:* distribution of mean scores by subject and by item. *Right:* mean score by condition (bars denote 95% confidence limits).

significantly better than chance ($t(27) = 6.6$, $p < .001$; see Fig. 2). In contrast, participants in the Scrambled condition preferred words over part-words on average 4.423 out of 8 ($t(25) = 1.389$, $p = .177$, n.s.). In addition, Varset scores were significantly better than Scrambled ($t(52) = 2.99$, $p = .004$). This difference was confirmed by a nonparametric Kruskal–Wallis rank sum test, $\chi^2 = 18.99, df = 1, p < .00013$.

Because effects that turn out to be significant in separate by-subject and by-item analyses may still be unreliable when all the random effects are considered jointly (Baayen, 2006; Jaeger, in press), we also fit a linear mixed model to the data using the lmer procedure (Bates, 2005).[5] A binomial logit-link linear mixed model fit to the scores yielded a significant effect of condition, $z = 3.323$, $p < .001$, confirming the outcome of the *t*-tests reported above.

Thus, participants failed to find words in unsegmented speech in the Scrambled condition, despite transitional probabilities supporting segmentation. At the same time, in the Varset condition, in which 20% of the sentences formed variation sets in the learning phase, participants performed significantly better, and better than chance.

The main effect of InOut was n.s. ($z = .158$, $p = .87$): in the Varset condition, both In words and Out words were learned reliably above chance ($p < .000001$ in both cases, Wilcoxon signed rank test with continuity correction), suggesting that the facilitatory effect of variation sets extended to the segmentation of words that did not participate in variation sets.

3. Experiment 2: learning phrase structure

Identifying word boundaries in continuous speech sets the stage for discovering higher order patterns such as phrase structure. In Experiment 2, we asked whether the presence of variation sets might facilitate the learning of phrasal constituents. Specifically, we tested whether identifying phrasal groupings defined by a phrase-structure grammar improves when the partial lexical overlap between adjacent sentences is consistent with constituency structure. This in turn might facilitate the learning of other syntactic patterns.

3.1. Participants and materials

Twenty-nine Cornell students participated for $6. In the learning phase, they listened to sentences containing pseudo-words. Sentences were generated by a set of phrase structure rules (Table 4). The resulting language consisted of sentences with two or three phrases, with Phrase3 being optional and appearing either in first position or at the end of the sentence. Phrase1 contained either two words from categories A and B, or a substituting word g1 (as in Experiment 1, capitals stand for lexical categories). Sentences were generated according to the probabilities indicated

[5] In addition to offering a more reliable picture of the data by accommodating crossed subject and item random effects, lmer tolerates unbalanced data (as when the numbers of subjects per condition differ), and also allows one to specify a distribution other than normal, as in our case. For the same reasons of non-normality, we also reported non-parametric tests throughout the analyses.

Table 4
Materials in Experiment 2.

Structure	Lexicon	Actual words
S1 (.70) → Phrase1 Phrase2 Phrase3	A = {a1, a2, a3}	`dro,kuhl,kleep`
S2 (.25) → Phrase1 Phrase2	B = {b1, b2, b3}	`goz,larp,cree`
S3 (.05) → Phrase3 Phrase1 Phrase2	C = {c1, c2, c3}	`bim,heeb,tood`
Phrase1 (.92) → A B	D = {d1, d2, d3}	`skiv,irg,bim`
Phrase1 (.08) → G	E = {e1, e2, e3}	`nerk,plam,quive`
Phrase2 (1.0) → C D	F = {f1, f2, f3}	`roo,tiv,yent`
Phrase3 (1.0) → E F	G = {g1}	`arv`

Table 5
A sample of four adjacent sentences presented during the learning phase in Experiment 2. While the abstract phrase structure and actual sentences are the same across Scrambled and Varset conditions, 20% of sentences in the Varset conditions were arranged such that a lexical phrase overlapped across two consecutive sentences.

Phrase structure	Scrambled condition	Varset condition
G C D	`arv bim skiv`	`arv bim skiv`
A B C D	`dro goz heeb irg`	`dro goz bim skiv`
A B C D E F	`kuhl larp tood bim plam roo`	`kuhl larp bim skiv plam roo`
E F A B C D	`nerk tiv dro cree heeb irg`	`quive tiv dro cree bim skiv`

in parentheses next to the each rule. Sentence length ranged from 3 to 6 words. No feature of individual words other than their distribution in the sentences signaled their class membership. The placeholders a1, a2,...,g1 were randomly assigned to lexical items for each participant. Nineteen monosyllabic words (Table 4) were recorded by a trained female voice.

For the learning phase, we arranged 365 unique sentences differently in Scrambled and Varset conditions. As in Experiment 1, no adjacent sentences in the Scrambled condition shared the same lexical item. In the Varset condition, 20% of sentences contained at least one overlapping phrase (Table 5).

There were 10 blocks of sentences arranged in variation sets (6 sentences each) interleaved with 10 blocks of sentences arranged in scrambled order. Any two adjacent sentences within a variation set overlapped partially, with at least one phrase in common. An English example would be:

– Push them **to school.**
– Take them **to school.**

Words were separated by 300 ms pauses and sentences by 750 ms pauses. Because all words within a sentence were separated by the same pause length and had a flat prosody, no acoustic feature signaled the presence of phrase boundaries.

The research question was whether or not participants in the Varset condition would exploit variation sets to group words into phrasal constituents, i.e., whether they would judge a 'C D' pairing more likely than a 'D E' pairing. Consequently, the test phase was a forced-choice task consisting of 12 trials, four for each phrase type ('A B', 'C D', 'E F'). A trial presented two pairs of words, one phrase pair (e.g., 'C D') and one pair that was a legal sequence in the language but straddled a phrase boundary (e.g., 'D E').

Frequencies and transitional probabilities of adjacent test words were as follows. For phrasal groupings: 'A B': 278 occurrences, $TP(B|A) = .53$; 'C D': 327, $TP = 1$; 'E F': 263, $TP = .87$. For groupings straddling phrases: 'B C': 278 occurrences, $TP(C|B) = 1$; 'D E': 224, $TP = .7$).

As in Experiment 1, an analysis of edit distances between successive sentences in the training data in the Varset condition (Fig. 3) revealed that in most variation sets,

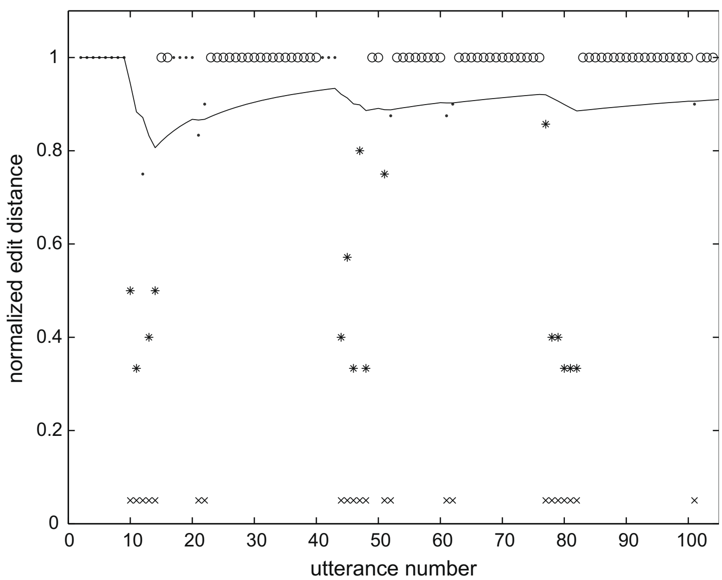

Fig. 3. The first 100 edit distances between successive utterances in Experiment 2 (for the legend, see Fig. 1). As in Experiment 1, the cumulative statistics of the edit distance values indicate that the alignments, where they exist, are significant, with a few exceptions.

the edit distance between the two sentences was significantly smaller than the baseline provided by the cumulative average.

3.2. Procedure

Participants were randomly assigned to either the Varset or Scrambled condition. They were asked to individuate the basic phrasal groupings of new sentences in a miniature language. As an example, the English sentence "My brother plays Nintendo at night" was described as having the following groupings: "(My brother) (plays Nintendo) (at night)." Learning lasted 18 min. At test, participants chose the pairs of words that they thought were more likely to be a group in the language.

3.3. Results

Participants in the Varset condition preferred phrases over part-phrases on the average 9.07 times out of 12, which was significantly better than chance ($t(14) = 6.35$, $p < .001$; see Fig. 4). Participants in the Scrambled condition preferred phrases over part-phrases on the average 7.36 times out of 12, which was also better than chance ($t(13) = 3.085$, $p < .01$). In addition, learning in the Varset condition was significantly better than in the Scrambled condition ($t(27) = 2.60$, $p < .015$; Kruskal–Wallis test, $\chi^2 = 16.37, p < .00052$). Thus, while learning did occur in both conditions, it was significantly better when variation sets were present in the learning phase. A binomial logit-link linear mixed model fit to the scores yielded a significant effect of Condition, $z = 2.672$, $p < .0075$, confirming this conclusion. The main effect of InOut was significant ($z = -2.243$, $p = .024$), suggesting that In phrasal units were learned better than Out phrasal units across conditions. However, both In and Out phrasal units were learned better than chance ($p < .0000001$, Wilcoxon signed rank test with continuity correction). Thus, variation sets facilitated learning of phrase structure even for lexical items that did not directly participate in variation sets.

4. Discussion

Variation sets are a pervasive phenomenon in natural child-directed speech. In a corpus analysis that prepared the ground for the present psycholinguistic study, we found that 58.6% of sentences in child-directed speech reside within variation sets. Thus, partial lexical overlap exists between a substantial portion of neighboring sentences. This suggests that child-directed speech is replete with local (hence, easy to use) cross-sentential cues to linguistic structure. By looking at language one sentence at a time, most of us in the past underestimated the amount of information available for learning (the work of Morgan, Meier, and Newport (1989), who studied the effects of cross-sentential cues in a controlled setting, is a notable exception). The full implications of the prevalence and composition of variation sets for developmentally-in-

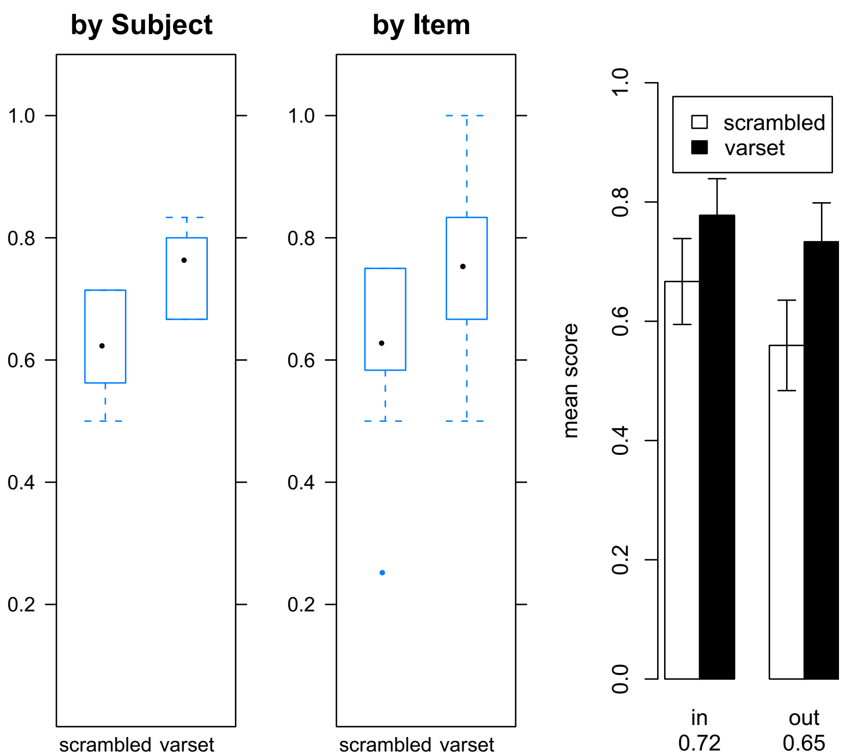

Fig. 4. Experiment 2. *Left:* distribution of mean scores by subject and by item. *Right:* mean score by condition (bars denote 95% confidence limits).

formed computational theorizing about language acquisition still await exploration (Brodsky, Waterfall, & Edelman, 2007).

The results of Experiment 1 and Experiment 2 indicate that arranging 20% of the learner's input in variation sets (corresponding to a conservative estimate from real child-directed speech that considers only open-class words) facilitates the discovery of linguistic structure at two different levels of analysis: finding words in continuous speech, and identifying phrasal constituents. Variation sets offer immediate and effective cues to linguistic structure by making it possible for the learner to resort to local (hence, computationally inexpensive), statistically sound discovery procedures based on alignment and comparison of successive utterances.

Current unsupervised computational approaches to finding structure typically are non-incremental, and rely on global cues – they amass statistical evidence over the entire learning experience (be it within an experimental session of 6 min or over a sample corpus of language) to infer the reliability of candidate structures. This is true both in lexical learning (e.g., Brent, 1999) and in syntax learning (e.g., Solan, Horn, Ruppin, & Edelman, 2005). This makes global approaches computationally costly (e.g., a word learning algorithm must maintain all possible candidate segmentations), as well as cognitively implausible.

Indeed, the lack of learning of our subjects in the Scrambled condition of Experiment 1 suggests that global alignment is not resorted to even for a small lexicon. Given the small lexicon in Experiment 1, spurious variation sets interleaved by one or two sentences were very likely to occur,[6] and yet subjects did not seem to have used such non-local alignments. In contrast, in the Varset condition, in which local alignment cues were present, learning did occur, even for words that did not participate in variation sets in training. Presumably, once lexical candidates are revealed in a variation set, they are also more recognizable when they occur in other sentences, thus promoting in turn the segmentation of novel words.

The results of our Experiment 2 may be compared to earlier work in artificial language learning that used cross-sentential cues, such as that of Morgan et al. (1989). These researchers found that when an artificial grammar was augmented with the substitution of phrases and variations in phrasal order, learning improved with respect to a baseline condition that (1) contained no such variations, and (2) whose adjacent sentences were merely repeated. The stimuli of Morgan et al. (1989), which included extra visual cues to category membership, consisted of pairs of aligned written sentences and geometrical figures on the screen. In contrast, in our experiments sentences were presented sequentially in their natural auditory modality and did not overlap in time or space.

More recently Thompson and Newport (2007) examined the effects on syntax learning of partially overlapping material between sentences, presented in the auditory modality. They did not, however, control variation sets as such; rather, they constructed increasingly more complex languages that gave rise to variation sets, and were able to show that more complex grammars could actually be easier to learn. In contrast, the present study is the first to manipulate only the order of presentation of the stimuli (variation sets), while maintaining the same complexity of the language across learning conditions (indeed, the very same sentences were used in both conditions for each experiment). This is, therefore, a direct demonstration that the language discovery procedures proposed by Harris (1946) can also contribute to language learning.

In summary, the positive effects of variation sets in our two experiments suggest that learners can reuse the same algorithmic building blocks – alignment, comparison, and, presumably, significance assessment – for different levels of linguistic structure (here, lexical and phrasal units). We are presently extending our approach to investigate whether variation sets also facilitate the learning of other core features of language, such as lexical categorization, long-distance dependencies, and recursion.

[6] When gaps of one utterance were allowed, the percentage of utterances in variation sets was 98% (Varset condition) and 100% (Scrambled) in the first experiment and 99.7% (Varset condition) and 45% (Scrambled) in the second experiment.

Acknowledgement

This study was realized while L. Onnis was at Cornell University, and was supported by Grant 5R03HD051671-02 from NICHD to L. Onnis, and by grant NIH Grant PO1 HD40605 to Susan Goldin-Meadow and Janellen Huttenlocher at the University of Chicago.

References

Baayen, R. H. (2006). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
Bates, D. (2005). Fitting linear mixed models in R. *R News, 5*, 27–30.
Brent, M. R. (1999). An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning, 34*, 71–105.
Brodsky, P., Waterfall, H. R., & Edelman, S. (2007). Characterizing Motherese: on the computational structure of child-directed language. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th cognitive science society conference*. Austin, TX: Cognitive Science Society.
Dutoit, T. (1997). *An introduction to text-to-speech synthesis*. Dordrecht: Kluwer.
Edelman, S., & Waterfall, H. R. (2007). Behavioral and computational aspects of language and its acquisition. *Physics of Life Reviews, 4*, 253–277.
Harris, Z. S. (1946). From morpheme to utterance. *Language, 22*, 161–183.
Hoff-Ginsberg, E. (1986). Function and structure in maternal speech: Their relation to the child's development of syntax. *Developmental Psychology, 4*, 155–163.
Jaeger, F. T. (in press). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models, *Journal of Memory and Language*.
Küntay, A., & Slobin, D. (1996). Listening to a Turkish mother: Some puzzles for acquisition. In D. Slobin & J. Gerhardt (Eds.), *Social interaction, social context, and language: Essays in honor of Susan Ervin-Tripp* (pp. 265–286). Hillsdale, NJ: Lawrence Erlbaum Associates.
Morgan, J. L., Meier, R. P., & Newport, E. L. (1989). Facilitating the acquisition of syntax with cross-sentential cues to phrase structure. *Journal of Memory and Language, 28*, 360–374.
Newport, E. L., Gleitman, H., & Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and noneffects of maternal speech style. In C. Snow & C. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 109–150). Cambridge, MA: Cambridge University Press.
Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169–225.

Rowland, C. F., Pine, J. M., Lieven, E. V. M., & Theakston, A. L. (2005). The incidence of error in young children's wh-questions. *Journal of Speech, Language, and Hearing Research, 48*, 384–404.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Solan, Z., Horn, D., Ruppin, E., & Edelman, S. (2005). Unsupervised learning of natural languages. *Proceedings of the national academy of science, 102*, 11629–11634.

Szmrecsanyi, B. (2005). Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory, 1*, 113–149.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: the role of transitional probability. *Language Learning and Development, 3*, 1–42.

Waterfall, H. R. (2006). *A little change is a good thing: Feature theory, language acquisition and variation sets*. Ph.D. thesis, University of Chicago.

Waterfall, H. R. (submitted for publication-a). Relation of variation sets to noun and verb development.

Waterfall, H. R. (submitted for publication-b). Relation of variation sets to syntactic development.