Local statistical learning under cross-situational uncertainty

Luca Onnis (lucao@hawaii.edu)
Department of Second Language Studies, University of Hawaii, Honolulu, HI 96822 USA
Center for Second Language Research, University of Hawaii, Honolulu, HI 96822 USA

Shimon Edelman (edelman@cornell.edu)
Department of Psychology, Cornell University, Ithaca, NY, 14853, USA

Heidi Waterfall (heidi.waterfall@gmail.com)
Department of Psychology, Cornell University, Ithaca, NY, 14853, USA
Department of Psychology, University of Chicago, Chicago, IL 60637, USA

Abstract

Statistical learning research often assumes that learners collect global statistics across the entire set of stimuli they are exposed to. In naturalistic settings, this assumption of global access to training data is problematic because it implies that the cognitive system must keep track of an exponentially growing number of relations while determining which of those relations is significant. We investigated a more plausible assumption, namely that learning proceeds incrementally, using small windows of opportunity in which the relevant relations are assumed to hold over temporally contiguous objects or events. This local statistical learning hypothesis was tested on the learning of novel word-to-world mappings under conditions of uncertainty. Results suggest that temporal contiguity and contrast are effective in multimodal learning, and that the order of presentation of data can therefore make a significant difference.

Keywords: statistical learning; cross-situational learning; variation sets; language acquisition.

What principles guide learning from multiple parallel streams of sensory information? How can humans find structure in sustained exposure to auditory and visual stimuli? Much of learning can be characterized as finding patterns in space and time under conditions of high uncertainty – from deriving categories from experience (e.g., Tenenbaum & Griffiths, 2001), through learning word meanings from their co-occurrence with perceived events in the world (e.g., Frank, Goodman, & Tenenbaum, 2009), to acquiring the different levels of linguistic structure (e.g., Solan, Horn, Ruppin, & Edelman, 2005).

Behavioral studies in statistical learning (e.g., Gómez & Gerken, 2000) indicate that infants, children, and adults can extract regularities across a set of exemplars distributed in time and/or space. A core assumption of such studies is that to extract such regularities, learners collect global statistics across the entire set of stimuli they are exposed to, often over multiple trials or training sessions.

When applied to naturalistic learning, this 'global' assumption is problematic: it requires that the cognitive system keep track of an exponentially growing number of relations among various pieces of data while identifying the significant relations. A more plausible assumption, investigated here, is that learning proceeds incrementally, using small windows of opportunity in which the relevant relations are those that hold over spatially and/or temporally neighboring objects or events.

For example, in a study by Onnis, Waterfall, & Edelman (2008), adult learners were asked to individuate the novel words of an "alien" language from unsegmented sentences such as *kedmalburafuloropesai*. In the absence of acoustic and prosodic cues (sentences were generated by speech synthesis software), each sentence could in principle be composed of a range of possible words ranging from a single long word (as is not uncommon in polysynthetic languages such as West Greenlandic), to as many words as there were syllables. Onnis et al. (2008) found that learners were significantly better at the word segmentation task when some consecutive sentences in the training set overlapped in some of their syllables (e.g., *kedmalburafuloropesai* followed by *rafuloro*), compared to a control condition in which the order of the same set of sentences was scrambled so that no parts of adjacent sentences matched. When aligned, the partially matching sentences suggest candidate

units (here, *kedmalbu*, *rafuloro*, *pesai*) without the need for learners to entertain all possible candidates over a long series of sentences. Local partial repetitions across sentences thus facilitate learning. Importantly, the study also found evidence for a "trickle-down" learning effect: not only did learners more reliably prefer word units heard in pairs of partially matching sentences, but also other valid units that never occurred in such pairs (e.g., *gianaber*, *kiciorudanamjeisulcaz*).

While the Onnis et al. (2008) study provided evidence for contingency and proximity as basic principles of structure inference from auditory cues, it did not investigate the alignment of linguistic structures with perceptual data from other modalities such as vision. Typically, there are many potential word-to-world mappings, and the learner must solve this correspondence problem in the face of uncertainty (Quine, 1980).

Can local learning based on principles of alignment and comparison reduce the combinatorial explosion of hypotheses arising in the process of matching words to their referents? Although statistical regularities alone can drive learning of word reference (e.g., Yu & Smith, 2007), the contribution of partial matches of adjacent stimuli to such learning was unknown.

The experiments described below were modelled after Yu & Smith's (2007) cross-situational word learning paradigm in adults. In each learning trial, participants saw multiple novel pictures and simultaneously heard multiple novel words, creating ambiguity regarding correct word-to-picture mappings. For instance, if four words and four pictures were presented on a single trial, there could be 4 x 4 = 16 possible word-referent combinations. The participants' task was to infer word-picture mappings across these training trials. At test, participants heard a single novel word and had to select one picture out of four that went with that word. Importantly, the subjects' ability to learn in this task indicates that they tracked the relations between words and referents across multiple trials, hence the term cross-situational learning. Other studies have shown that children can also make cross-situation comparisons when learning novel word-world pairings (e.g., Akhtar & Montague, 1999; Childers & Paik, 2009).

How exactly are the ambiguities in such learning tasks resolved? According to Yu & Smith (2007), cross-situational learning works as follows. Consider a simple learning scenario that consists of 4 trials as in Table 1, with each trial ambiguously presenting two words (in lower-case) and two referents (in upper-case). In trial 1, learners could mistakenly link the word a to the referent B; later they could successfully rectify their mistake and discover the correct a-A mapping, if all the following conditions applied: 1) they registered that word a occurred in trial 4 without the incorrect referent B; 2) they remembered the prior word-referent pairing; 3) they registered both co-occurrences and non co-occurrences among all possible pairings; and 4) they calculated the right statistics.

Table 1

Trial	Words	Potential referents in scene
1	**a** b	B **A**
2	c d	D C
3	e f	E F
4	g **a**	G **A**

Table 2

Trial	Words	Potential referents in scene
1	**a** b	B **A**
2	g **a**	G **A**
3	c d	D C
4	e f	E F

Clearly, the proposed mechanism is capable of tracking multiple statistical relations both locally (over adjacent trials) and *globally* (over the entire learning experience). For the purposes of this study, we refer to this mechanism as a *global statistical learner*.

There are reasons to believe that this mechanism may be unrealistic, especially when scaled up to naturalistic learning situations. First, it would be difficult for learners to keep track of all possible pairings while noting which did and which did not occur at any given time. Indeed, there is evidence that learners often fail to track the non-occurrence of two elements in artificial language experiments (e.g., Smith, 1966). Second, the results of Onnis et al. (2008) suggest that local alignment and comparison aid the discovery of structural relations in artificial language learning, but the

global statistical learner makes no special use of local cues. The hypothesized distinction is supported by Smith & Yu's (2008) finding that some word-picture correspondences were learned better than others when the cross-situational paradigm was adapted to infants, leaving open the possibility that chance consecutive alignments in the uncontrolled presentation order of the stimuli were responsible for the differences.

We were able to confirm this hypothesis empirically: a bootstrap analysis of the training trials of Yu & Smith (2007) (see Experiment 1 below) showed that the mean percentage of consecutive trials that had partial overlapping word-reference pairs was 90.4% (SD=5.1%; the bootstrap procedure generated 100 randomized sequences of the given set of trials). We thus surmise that partially overlapping consecutive trials were likely to have been present in the randomly generated learning sequences of Yu & Smith (2007) and contributed to the disambiguation of the word-scene mappings. Such accidental local alignment may have played an important role in enabling fast mapping in both the adult and the infant studies of Yu & Smith.

To better understand the distinction between local and global accounts of statistical learning, let us compare the scenario described in Table 1 with a new learning scenario. In Table 2, the pair a-A occurs in trial 2 immediately after trial 1. In the Words column, there is a single word a that stays constant across two successive trials; likewise, in the Scenes column there is a single constant referent A across two successive scenes. Thus, by simultaneously keeping track of the two constant elements in Words and Scenes (the auditory and visual modality), the learner can identify a unique word-mapping that remains constant over any two successive trials, without the need to hold in memory other potential relations, as the global learning strategy would require.

These considerations allow us to make the following predictions. If, on the one hand, global cross-trial statistics accumulated over all learning trials is what drives cross-situational learning, then learning in scenarios 1 and 2 above should be equally successful (the global statistics are the same—the only difference is the order of presentation of the trials). If, on the other hand, local alignment is partly responsible for learning, the scenario depicted in Table 2 should lead to significantly better learning than the scenario depicted in Table 1. A stronger case could be made that, in the absence of partially overlapping trials in the learning phase, learning of word-referent mappings would not differ from chance.

Experiment 1

Method

Participants. Seventeen students at the University of Hawaii participated and were compensated $5 each.

Materials. Each trial in a set of 18 learning trials contained four spoken words and four pictures of individual objects, with no information about specific word-picture correspondence. The task was to learn nine word-referent pairs, each repeated eight times, over the 18 training trials. This 4 (auditory) x 4 (visual) learning scenario yields a large number of possible word-referent associations: 16 per trial. Because the referent of a word cannot be unambiguously determined during any single trial, disambiguation must rely on information from multiple trials.

Procedure. We assigned participants to one of two conditions: 1) a Contiguous condition, in which 80% of trials had one word-referent pair in common between two consecutive trials, and 2) a Scrambled condition, in which only 5% of consecutive trials contained auditory or visual overlap. At test, participants received nine test trials each with a single word from among those heard during training and had to choose the correct referent among four possibilities. Importantly, because the two conditions differed only in the order of trials, the global statistics of word/picture co-occurrence were identical. This allowed us to make differential predictions, as follows. On the *global statistical learning* account, learners solve the indeterminacy problem by keeping track of multiple word-referent statistics across many individually ambiguous words and scenes, possibly over the entire experiment (Yu & Smith, 2007). Thus, learning should not differ across our two conditions. Conversely, on the *local statistical learning* account (see also Onnis, Waterfall, & Edelman, 2008), learners benefit from the contiguous arrangement of partially overlapping trials and should therefore learn better in the Contiguous condition.

Results

Our results supported the local learning hypothesis: a one-way ANOVA revealed a main effect of Condition, $F(1,15)=5.4$, $p<.05$). Separate t-tests showed that participants in the Contiguous condition learned better than theoretical chance

(2.25), M=3.67, SD=1.23, $t(8)$=2.86, p<.03. Conversely, learners in the Scrambled condition failed to learn above chance, M=2.5, SD= 0.75, $t(7)$=.00, p=1.0. A nonparametric Kruskal-Wallis rank sum test yielded a marginal difference between the conditions (chi-squared = 2.81, df = 1, p < 0.09). A mixed model with Subject and Item as random effects (R procedure lmer, with a binomial link function) yielded a similar marginal advantage for the Contiguous condition (z=1.703, p < 0.09).

Contrary to an exclusively global account of statistical learning, this pattern of results suggests that learning proceeded on a more local trial-by-trial basis, exploiting temporally contiguous relations.

Experiment 2

Experiment 1 suggests that in the absence of local statistics in the form of partial self-repetitions learners found it much harder to discover the correct word-reference pairings. However, the specifics of how presentation order affected learning in the Contiguous condition remained unknown. For instance, it is possible that learning occurred in the last two trials, driving this effect in its entirety (i.e., learners could just be remembering the last few trials, and if they happened to contain repetition, that is what was learned). In Experiment 2, we thus divided the word-referent items into two categories; for one, the repetitions always happened in the first half of training, and for the other – in the second half. If learning is indeed driven by a 'recency' effect, items in the second half should be learned better than those in the first half.

Method

Participants. We recruited 36 students at the University of Hawaii who had not participated in Experiment 1. Each received $5.

Materials. The same training and test materials as Experiment 1 were used. The only difference was that four of the word-referent pairings occurred across overlapping trials in the first half of the training set (i.e., in the first 9 trials, Early Pairs), while another four occurred in the last 9 trials (Late Pairs). There was no Scrambled condition.

Procedure. The same procedure as Experiment 1 was used.

Results

A one-way ANOVA revealed a main effect, $F(1,17)$=101.2, p<.0001), thus replicating the positive effect of contiguous trials on learning obtained in Experiment 1. Furthermore, there was no effect of Order between Early and Late Pairs, $F(1,17)$=0.596, p=.451. This was confirmed by Kruskal-Wallis and mixed-effect tests. Separate t-tests showed that both Early and Late Pairs were learned better than chance (Early Pairs, M= 1.94, SD=.93, $t(17)$=4.27, p<.001; Late Pairs, M= 2.22, SD=1.35, $t(17)$=3.83, p<.01), although the mean for Late Pairs was numerically higher. Thus learning did not appear to be driven by a 'recency' effect: subjects learned equally well items presented in either halves of training.

Experiment 3

How dependent are the effects found in Experiments 1 and 2 on the degree of uncertainty in the training data? In Experiment 3, we raised the level of cross-situational uncertainty by increasing the number of word-referent pairs to be learned from 9 to 18, while keeping the same within-trial ambiguity as in the previous experiments (4 x 4 =16 possible mappings in each individual trial), and reducing the number of repetitions of the individual word-referent pairs with respect to Experiments 1 and 2. Learners were assigned either to a Contiguous condition or a Scrambled condition, as in Experiment 1.

Two sets of predictions were made. For a global statistical learner that collects statistics over the entire set, the 18 word condition should lead to better learning than in the 9 word conditions of Experiment 1 and 2. Yu and Smith (2007) argued that learning more word-referent pairs should be easier for the global statistical learner because there would be fewer spurious pairings and thus, those pairings that do occur would be more systematic. The prediction for a local statistical learner is that – although the number of words and referents to be tracked increases – it should still be easier to learn 18 word-referent pairs in the Contiguous condition, where partial self-

repetitions can more immediately winnow out the correct pairs.

Method

Participants. We recruited 31 students at the University of Hawaii who had not participated in Experiment 1 or 2. Each received $5.

Materials. We constructed a set of 18 word-referent pairs to be learned over 26 learning trials. As in Experiments 1 and 2, a trial contained four spoken words and four pictures of individual objects, with no information about specific word-picture correspondence.

Procedure. The same procedure as Experiments 1 and 2 was used. Because there were 18 pairs to be learned, the Test phase comprised 18 test trials.

Results

A one-way ANOVA revealed no effect of Condition, $F(1,29)=1.134$, $p=.29$), suggesting that learning occurred both in the Scrambled and Contiguous conditions. This was confirmed by separate t-tests: participants in the Contiguous condition learned better than chance (M= 9.25, SD=3.57, $t(15)=5.325$, $p<.001$, chance level=4.5). Learners in Scrambled also learned above chance (M=8, SD= 2.9, $t(15)=4.669$, $p<.001$), although the mean in Contiguous was numerically higher than in Scrambled. Notice that the means in Experiment 3 are considerably higher than those in Experiments 1 and 2, replicating Yu and Smith's (2007) finding that a larger lexicon is actually more manageable to learn than a small one in cross-situational learning.

Taken together, these data suggest that a global statistical learning account cannot be entirely ruled out, consistent with the view that with more pairs to be learned, the number of spurious relations diminishes, thus helping global learners to reduce cross-situational uncertainty. However, we wanted to explore the data more thoroughly to investigate the patterns of learning, both by subjects and by items. We thus computed the grand mean M and standard deviation S of learning performance over all subjects, and treated data from each subject whose personal mean m was too far below the grand mean as an outlier.

With the bound for outlier detection set to 1.5 S below the grand mean M ($m < M - 1.5$ S), the difference between Scrambled and Contiguous was not significant. However, when the lower bound was set to 1.0 S below the grand mean M ($m < M - 1.0$ S), data from the 25 (out of 31) participants whose learning performance exceeded the threshold revealed a significant effect of Condition. The Kruskal-Wallis test yielded a significant difference between conditions (chi-squared = 5.94, df = 1, $p < 0.01$). A mixed-effects model (same procedure as in Exp.1) also showed a significant effect of Condition ($z = 2.30$, $p < 0.02$).

This pattern of findings can be interpreted as follows: the presence of partially overlapping trials helped good learners, but not poor ones. Indeed, it may be that what makes a good learner is, in part, the ability to use local information immediately available in the input (more on this in the Discussion section).

Finally, we asked whether the effects of partial self-repetitions in the Contiguous condition are confined to the specific items that enjoy the special distributional environment, or whether the training regimen has more general effects. During training, half of the word-referent pairs were presented in contiguous partial-self repetitions (IN pairs), while the other half were not (OUT pairs; as an example drawn from Table 2, the pair A-a would be an IN pair, while C-c would be an OUT pairs across the learning phase). A one-way ANOVA with Item Status as factor (IN vs. OUT) revealed no effect of Item Status, $F(1,15)=1.552$, $p=.23$). Both pair types were learned above chance (IN Pairs, M= 4.25, SD=1.98, $t(15)=4.54$, $p<.001$; OUT Pairs, M= 5, SD=2.31, $t(15)=5.2$, $p<.001$, chance level=2).

Discussion

Under natural circumstances the joint presentation of a word and a scene offers many possible word-object pairings (Quine, 1960). To explore the potential benefits of temporally local learning in such a situation, we replicated a study by Yu and Smith (2007), and investigated whether local learning yields global benefits that extend beyond the relations encountered locally, thus helping the learner manage the computational complexity of structural inference.

Yu and Smith (2007) and Smith and Yu (2008) proposed that both adults and children solve the

problem of word-to-world mapping by keeping track of multiple word-referent statistics across many individually ambiguous words and scenes. On this account, a learner could mistakenly link a word to a referent, but correct the mistake by, first, registering on a subsequent trial that the word had occurred without the earlier, wrong referent; second, by remembering the prior word-referent pairing; and, third, by registering both co-occurrences and non co-occurrences. Our null result in the Scrambled condition suggests that such global statistical accounting may be beyond learners' capabilities. In comparison, the learners' success in the Contiguous condition suggests two possible cues that may have helped the subjects: first, the reappearance of a word and its referent and second, the fact that on the very next trial, everything changed *except* that particular word-referent mapping.

Yu and Smith (2007) reported learning that was significantly better than chance. How can we account for their finding? The order of presentation of learning trials in their experiment was random (the experiment otherwise contained the same type and number of stimuli as ours). We generated 100 separate randomizations of those trials and found that the number of partially overlapping contiguous trials was very high (M=90.4%, SD=5.1%). Thus, the learning regime of Yu and Smith is more similar to our Contiguous than to our Scrambled condition.

Taken together, our results suggest that encountering a consistent pairing of words and referents in a temporally contiguous manner facilitates learning, compared to a randomly scrambled presentation of the very same stimuli pairs. Moreover, the advantage of contiguous presentation seems to be lost on poor learners. Whether this difference is due to working memory limitations, inattention, or some other factors would need to be investigated in the future.

Contiguity and contrast were first invoked by Aristotle (De mem. et rem.) as fundamental laws of association. Following the early insights of Hume, researchers have come to appreciate the crucial role of statistical inference in ensuring the reliability of experiential learning. In this paper, we considered the task of establishing word-referent association using statistical patterns of experience. A recently proposed theoretical framework, ACCESS (Align Candidates, Compare, Evaluate Statistical/Social Significance) aims to explain the learning of structure in space and time in terms of general principles of cognitive computation (Goldstein et al., 2010). In agreement with those principles, our results suggest the effectiveness of temporal contiguity and contrast in multimodal learning under conditions of uncertainty, and the importance of order of presentation of learning materials – a finding that has intriguing implications for various practical learning situations.

References

Akhtar, N., & Montague, L. (1999). Early lexical acquisition: The role of cross-situational learning. *First Language*, 19, 347-358.

Aristotle, *De Memoria et Reminiscentia*. On Memory and Recollection. Available at http://classics.mit.edu/ Aristotle/memory.html.

Childers, J., & Paik, J.H. (2009). Korean- and English-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language*, 36, 201-224.

Frank, M.C., Goodman, N.D., and Tenenbaum, J. B. (2009), Using speakers' referential intentions to model early cross-situational word learning, *Psychological Science*, *20*, 578-585.

Goldstein, M.H., Waterfall, H.R., Lotem, A. Halpern, J., Schwade, J., Onnis, L., and Edelman, S. (2010). General cognitive principles for learning structure in time and space, *Trends in Cognitive Sciences*, 14, 249-258.

Gómez, R.L., & Gerken, L.A. (2000) Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*, 178-86.

Onnis, L., Waterfall, H. & Edelman, S. (2008). Learn locally, act globally: Learning language from variation set cues. *Cognition*, *109*, 423-430.

Quine, W. V. O. (1960). *Word and Object*. Cambridge, MA: MIT Press.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558–1568.

Solan, Z., Horn, D., Ruppin, E. and Edelman, S. (2005). Unsupervised Learning of Natural Languages, *PNAS*, *102*, 11629-11634.

Tenenbaum, J.B., and Griffiths, T.L. (2001). Generalization, similarity, and Bayesian inference, *Behavioral and Brain Sciences*, *24*, 629-641.

Yu, C., and Smith, L. B. (2007) Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(*5*), 414-420.