Adults Track Multiple Hypotheses Simultaneously during Word Learning

Suzanne Aussems (s.aussems@warwick.ac.uk)
Department of Psychology, University of Warwick
University Road, Coventry CV4 7AL, United Kingdom

Paul Vogt (p.a.vogt@uvt.nl)
Tilburg center for Cognition and Communication, Tilburg University
P.O. Box 90153, 5000 LE Tilburg, the Netherlands

Abstract

Cross-situational learning is a basic mechanism that enables people to infer the correct referent for a novel word by tracking multiple hypotheses simultaneously across exposures. Previous research has shown that adults are capable of exploiting cross-situational information, but recently this gradual statistical learning mechanism has been put under debate by researchers who argue that people learn via a fast mapping procedure. We compared the performance of adult participants on a word learning task in which information was manipulated cross-situationally with the performance of simulated learning strategies. Experimental evidence indicates that adults use cross-situational learning, which appears to be a robust mechanism that facilitates word learning even under cognitively demanding circumstances.

Keywords: Cross-situational learning; artificial word learning; propose-but-verify strategy; fast mapping.

Introduction

When learning the meaning of novel words, a learner has to deal with the indeterminacy of word meaning. Quine (1960) introduced the example of the word *gavagai* uttered by a native speaker of Arunta upon seeing a rabbit. If an anthropologist wants to know the meaning of the word, he has to hypothesize about the correct referent for the word. It could mean *rabbit*, *furry*, *rabbit tail*, or even something utterly different than that. Quine (1960) noted that, for any given example, there will be many hypotheses consistent with the meaning of the word. The following question then arises: How do we infer the correct hypothesis about the meaning of a novel word? Presumably, we could consider all possible mappings, examine multiple situations, and accept only those mappings consistent with all, or at least most, of the different situations encountered. This basic mechanism is called “cross-situational learning” (Siskind, 1996).

Over the past few decades, researchers have shown that adults can and, indeed, do learn novel words using cross-situational information (Gillette, Gleitman, Gleitman, & Lederer, 1999; Yu & Smith, 2007; Piccin & Waxman, 2007; Smith, Smith, & Blythe, 2009, 2011). A cross-situational learner collects information across exposures to infer the correct referent for a target word from multiple hypotheses simultaneously. Recently, cross-situational learning (henceforth XSL) has been put under debate by researchers who believe that humans learn novel words using a *propose-but-verify* strategy (Trueswell, Medina, Hafri, & Gleitman, 2013).

Trueswell et al. (2013) conducted an experiment in which participants were repeatedly exposed to target words, whose meanings could be inferred from cross-situational information. Based on their results, however, they argue that people quickly propose one random hypothesis as to what the meaning of a target word is, and stick to this hypothesis as long as possible. If in a new situation the hypothesized meaning is no longer present, they propose another random hypothesis without any recollection of potential referents from previous situations, and this process is then repeated. Therefore, Trueswell et al. (2013) refer to this propose-but-verify strategy as a fast mapping procedure rather than a gradual statistical one. As such, their study contradicts earlier studies that pointed toward a XSL account for word learning.

We argue that a flaw in the experimental paradigm by Trueswell and colleagues may have induced the propose-but-verify strategy. In their experiments, as well as in many other word learning experiments (e.g., Smith et al., 2011), participants are subjected to a forced-choice paradigm in which they are instructed to click on the referent they think corresponds to a target word. As a consequence, participants have to propose a word-referent mapping on first exposures to target words. However, under natural circumstances, humans are not forced to decide on the meaning of a novel word on its first encounter; they may consider multiple hypotheses until they have obtained sufficient information across situations to make a decision. A possible solution for this problem is to use a look-and-listen paradigm in which the participants' gaze behavior provides an indication of their preference for an object, or multiple objects, when they hear a target word during the word learning procedure. Trueswell et al. (2013) also measured the eye movements of adult learners, but we suspect that the forced-choice paradigm they used on top of the eye-tracking paradigm might have biased their learning behavior. A follow-up study by Koehne, Trueswell, and Gleitman (2013) confirms our suspicion that a forced-choice paradigm might affect the learning strategies people use, but it remains unclear whether learners use a propose-but-verify strategy or XSL. In order to further investigate this, we use only a look-and-listen paradigm, which allows participants to consider multiple hypotheses during the word learning procedure without making a forced choice.

Trueswell et al. (2013) argue that multiple-hypothesis tracking may only occur under greatly simplified conditions in which learners are exposed to long consecutive sequences of learning instances for target words. However, Smith et al. (2011) found that when learning instances for target words were interleaved with learning instances for other target words, this only moderately affected learning success and learning speed, suggesting that human learners were able to use XSL even under these cognitively demanding circumstances. Taking into account these contrasting claims with regard to the influence of cognitive load on word learning strategies, we include a consecutive and interleaved condition in our experiment.

We present adult participants with a limited number of learning instances for each word to accentuate the differences between the performances of the two learning strategies under consideration. As a limited number of learning instances should lead to less successful learning with the propose-but-verify strategy as opposed to XSL, this allows us to investigate the extent to which the word learning performance of adult participants can be attributed to either one of these strategies. At the same time, the learning task becomes more difficult with a limited number of learning instances, especially with an interleaved presentation, which allows us to say something about learning under cognitively demanding circumstances. Using computer simulations, we set baseline measurements for different degrees of XSL and compare these with the word learning performance of adult participants. We expect that learners use XSL, regardless of whether they are presented with consecutive learning instances for target words, or learning instances that are interleaved with learning instances for other target words. Before presenting our study, we provide some more background on the considered learning strategies.

Learning Strategies

Cross-Situational Learning

Figure 1 aids in explaining a XSL procedure. Suppose a learner encounters the the word *timilo* for the first time. When the utterance is accompanied by the objects in Situation 1, a learner can hypothesize that each of these referents could refer to *timilo*. When the learner encounters the utterance again in Situation 2, he can verify or falsify his previously proposed hypotheses based on a change in the set of objects. One object from Situation 1 does not occur in Situation 2, which he can eliminate from the set of likely candidates for the target word-referent mapping. Moreover, one of the objects in Situation 2 was not present in Situation 1, which he can perceive as an unlikely candidate. When *timilo* is uttered in Situation 3, the learner can observe that again one object from the previous two situations can be eliminated, which was replaced by a new object that is unlikely to refer to the utterance since it was not present before. Finally, the learner can repeat this process in Situation 4 and eliminate the object that occurred in all previous situations, but not in this one. At this point, only the upper left object is consistent across all situations, making it the only possible referent for the word *timilo*. Thus, the learner can infer the correct word-referent mapping through a careful process of simultaneously proposing and verifying multiple hypotheses by eliminating referents across situations.

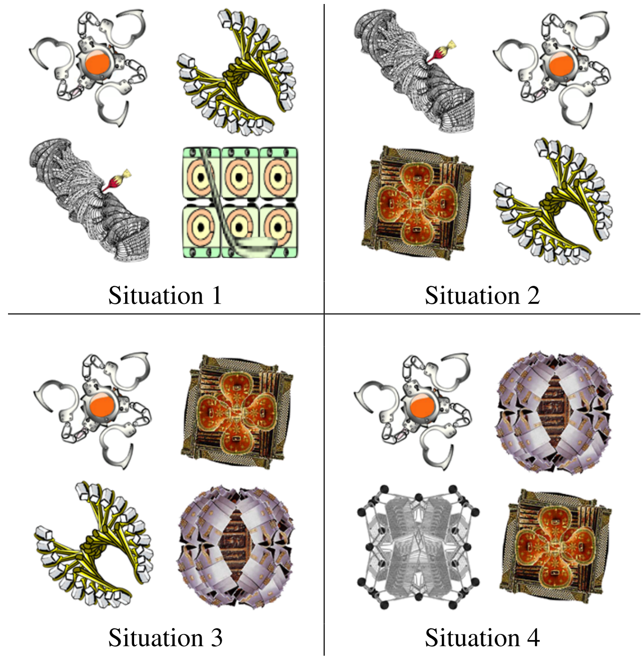

Figure 1: Four situations that display novel objects for the word *timilo*. Only the upper left object in Situation 1 is consistent across all situations. Stimuli are developed by Smith et al. (2011).

Propose-But-Verify Learning

In the propose-but-verify strategy (Trueswell et al., 2013), the learner will, in Situation 1, select one object at random as the proposed referent. If this object is still present in Situation 2, he will select it again. If not, he will propose another object that is present in Situation 2. Trueswell et al. (2013) argue that the learner does so without keeping track of previous exposures, so this alternative object is selected at random and could well be the newly introduced object. This same procedure continues until the experiment stops. So, although after four exposures a perfect cross-situational learner will have inferred the correct mapping, the propose-but-very strategy does not guarantee successful learning; not even after many more exposures. However, the probability that the correct mapping is learned increases with the number of exposures if the target referent is consistently present in all situations.

Baseline Measurements

Smith et al. (2009) suggest that in order to test if word learners do indeed use XSL, their performance should be compared with the performance of the best possible non-cross-situational learner. A baseline performance at chance-level, which was used in the study by Yu and Smith (2007), would not suffice because this baseline can be outperformed by another non-cross-situational learner; a learner that uses a single one of the training exposures to reduce the set of potential referents on test trials. This so-called *one-exposure learner* does not integrate information from multiple situations. In order to demonstrate that humans use cross-situational information for word learning, it must be demonstrated that they outperform the one-exposure learner.

With regard to more recent developments presented by Trueswell et al. (2013), we argue that testing against a one-exposure learning baseline does not conclusively demonstrate XSL, as Smith et al. (2009) proposed. Comparing the performance of participants with the performance of the one-exposure learner indeed demonstrates that participants integrate information across exposures, but not necessarily that they track multiple hypotheses simultaneously to infer word-meaning mappings, which is also a key characteristic of XSL. To demonstrate that participants truly use XSL, we must also show that they outperform the propose-but-verify learner by Trueswell et al. (2013). As a single hypothesis-based strategy, the propose-but-verify strategy does not qualify as a fully developed XSL strategy. Thus, comparing the learning performance of our participants with the performance of a propose-but-verify learner provides a second baseline measurement for XSL performance; if this baseline is exceeded, then participants must be tracking multiple hypotheses simultaneously to infer word-meaning mappings.

Method

Design

We adapted the experiment of Smith et al. (2011) by reducing the number of training trials for each word from 12 to 5 and by monitoring performance using an eye-tracker instead of a forced-choice paradigm. The experiment used a within-subjects design in which the order of learning instances (consecutive versus interleaved) was counterbalanced across participants.

Participants

We collected data from 92 Dutch native speakers (47 females), all students in the Tilburg School of Humanities. Participants received course credit for participation. Ten participants were excluded from the analysis, because the eye tracker did not record their eye movements properly. The final sample consisted of 82 participants (45 females) between the age of 18 and 30 years old ($M = 22.46, SD = 2.72$).

Materials

In order to control for the potential role of sound symbolism, syllables containing rounded and unrounded vowels following Dutch phonotactical rules were randomly concatenated into nonsense words. This resulted in the following linguistic stimuli: *toekie, boezie, noebee, voolee, wootie, nieloo, wiepoe, veegoo, reezoo*, which were randomly assigned to the practice procedure, consecutive block, or interleaved block. Audio samples of the words were generated using a Dutch online text-to-speech generator[1]. The length of the words was controlled for by keeping the number of letters and syllables the same.

[1] http://www.fluency.nl/international.htm

To create a visual stimuli set for the word learning procedure, nine sets of eight stimuli were randomly taken from the set of 120 pictures of nonsense objects developed by Smith et al. (2011). There was no overlap between sets to prevent participants from using mutual exclusivity (Markman & Wachtel, 1988). One target object was retrieved from each set and labeled with one of the generated nonsense words. The remaining 63 stimuli served as distractor objects.

Five learning instances were created for each word. Cross-situational information was manipulated by replacing one of the distractor objects systematically with on of the other distractor objects on each learning instance. This allowed for the verification and falsification of potential word-referent mappings across learning instances.

In the consecutive block, full sets of five learning instances for the target words were randomly displayed. In the interleaved block, all first learning instances for the four words were randomly displayed, followed by all second, third, fourth, and fifth learning instances. At the end of each block, test trials were randomly displayed. During these test trials, participants were presented with the unique set of eight potential referents for each word on a 2x4 display.

Apparatus

A SMI Vision RED 250 remote eye tracking system was used for stimuli presentation and data collection. Stimuli were presented on a 22" computer screen via SMI Experiment Center 3.3 and simultaneously eye gaze data from the eye tracker were collected via SMI iView X. Bright lights on both sides of the computer screen provided optimal calibration.

Procedure

Participants were tested individually in a soundproof booth in the lab. The distance between the participants and the computer screen was approximately 70 cm. The eye-tracker was calibrated for each participant using a 9-point calibration. The experimenter validated if the estimation of the eye position was indeed close to the known calibration points. If errors occurred, the calibration session was repeated. After the experimenter left the booth, participants put on headphones and started the experiment. They were instructed via the screen to try and map a target word to its correct referent, with the hint that the correct referent was always displayed in the context of the utterance. Note that participants were not explicitly instructed to use a particular learning strategy.

Each participant completed a training and test phase for one practice word, and subsequently, for a consecutive and interleaved block of four words each. Participants either completed the consecutive block or the interleaved block first. During training, participants were exposed to four objects for 5000 ms whilst they heard the utterance of a target word. Test procedures for the consecutive and interleaved block followed when training within a block finished. During testing, participants were presented with the unique set of objects for 8000 ms and they were specifically instructed to look at the object that they thought corresponded to the uttered target word.

Data Analysis

Eye movements of the participants were analyzed using an Areas of Interest (AOI) approach. Equally-sized AOIs were drawn around all objects on training and test trials by a human coder. Participants' calibrations were corrected manually using the software and procedure by Cozijn (2006).

Looking times for each object were normalized by calculating the percentage for the amount of time a participant spent looking at each object on each trial. Learning speed was monitored by measuring the time participants spent looking at target objects during training. Learning performance was scored as the maximum number of words a participant learned during the test procedures of the consecutive block (range 0 to 4) and interleaved block (range 0 to 4). The object to which participants looked longer than 50% during each test trial was accepted as the chosen referent.[2] If this chosen object indeed corresponded with the target word the word-referent mapping was considered learned. The participants were scored one point for each learned word. If the chosen referent did not correspond with the played target word, or if none of the looking times met the $> 50\%$ threshold, participants received zero points.

Simulations

We simulated distributions of 82 performance scores for each discussed learning strategy; a number that corresponds to the number of participants we included in our analyses.

First, the one-exposure learner by Smith et al. (2009) was simulated by implementing the following algorithm:

1. For each word, choose one of the training trials at random and store the objects of these trials in memory.

2. Use the stored information on the testing trial by randomly selecting a referent from the set of objects that appeared on the stored training trial.

Second, the propose-but-verify account by Trueswell et al. (2013) was modeled using a simple algorithm that can be described as follows:

1. For each word, start by choosing the referent at random from the displayed context.

2. On any additional exposure to a word, remember the previous guess.

3. If the remembered guess is present in the current context select the referent; otherwise select a referent at random from the objects at display.

4. Select referents chosen on final training trials on test trials.

[2]The threshold of $> 50\%$ of the time spent looking at the object of preference was chosen, because this percentage indicates that participants could not have spent more time looking at any of the other objects. 13 out of 656 data points were corrected (1.98%).

Our computer simulations were configured in the most optimal manner: with perfect memory. The learning instances of the word learning procedure were manipulated such that information could be integrated cross-situationally to infer correct word-meaning mappings. Thus, simulating XSL would lead to a perfect performance.

Results

Experimental Findings

Table 1 shows the frequency distributions for the number of words learned in the consecutive and interleaved block of learning instances, showing that the majority of participants learned four words in the consecutive condition and three or four words in the interleaved condition.

Table 1: Frequency distributions of the number of words participants learned in the consecutive and interleaved block of the word learning procedure.

No. of words	No. of participants			
	Consecutive	%	Interleaved	%
0	1	1.2	1	1.2
1	1	1.2	13	15.9
2	10	12.2	22	26.8
3	22	26.8	22	26.8
4	48	58.6	24	29.3
Total	82	100.0	82	100.0

A paired-samples t-test showed that participants learned significantly more words in the consecutive block ($M = 3.40, SD = 0.84$) than in the interleaved block ($M = 2.67, SD = 0.12$), $t(81) = 5.58, p < .001, r^2 = .28$.

An independent t-test showed that participants who were presented with consecutive learning instances followed by interleaved learning instances ($M = 6.18, SD = 1.63$) did not perform significantly better than participants who received a reversed order of presentation modes ($M = 6.00, SD = 1.53$), $t(80) = 0.51, p = .618, r^2 = .06$.

Looking times for target objects during training were submitted to a 2 (presentation mode) x 4 (word) x 5 (learning instance) analysis of variance with a repeated-measures design. Mauchly's test showed that the assumption of sphericity had been violated for the main effect of learning instance, $\chi^2(9) = 114.84, p < .001$, and the interaction between presentation mode and learning instance, $\chi^2(9) = 41.49, p < .001$.[3]

Figure 2 shows the average amount of time participants spent looking at target objects on sets of five learning instances in each condition. The ANOVA revealed a main effect of presentation mode on the amount of time participants spent looking at target objects during training, $F(1,75) = 20.05, p < .001, \eta_p^2 = .21$. Participants spent more time looking at target objects in the consecutive condition ($M =$

[3]Huynh-Feldt corrections resulted in the exact same F-ratio's, p-values, and explained variances. We reported the uncorrected degrees of freedom for the sake of convenience.

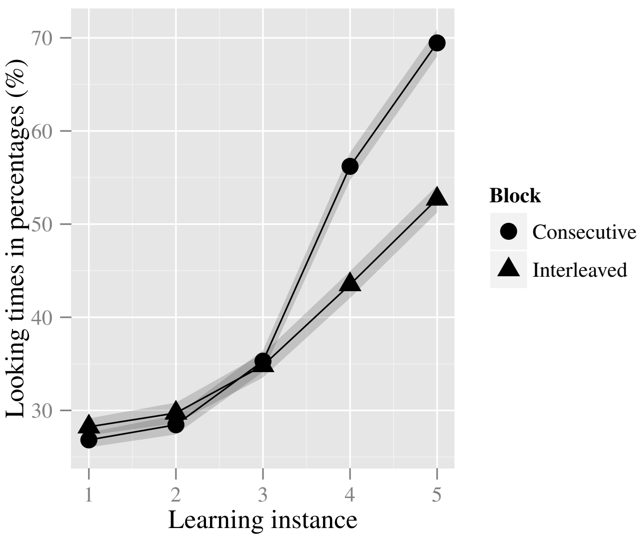

Figure 2: Looking times for target objects on learning instances in the consecutive and interleaved conditions. Shade represents standard error of the mean.

$43.08, SE = .92$) than in the interleaved condition ($M = 37.71, SE = 1.03$). The results also indicated a significant interaction between presentation mode and learning instance, on the amount of time participants spent looking at target objects on training trials, $F(4,300) = 30.37, p < .001, \eta_p^2 = .29$.

When learning instances were entered separately in the model for each condition, the ANOVA revealed no significant difference between first learning instances ($F(1,78) = 2.48, p = .119, \eta_p^2 = .03$), second learning instances ($F(1,80) = 0.91, p = .343, \eta_p^2 = .01$) and third learning instances ($F(1,81) = 0.10, p = .753, \eta_p^2 = .00$). However, there was a significant difference between fourth learning instances ($F(1,78) = 30.01, p < .001, \eta_p^2 = .29$) and fifth learning instances ($F(1,79) = 56.70, p < .001, \eta_p^2 = .42$), which indicated that participants looked longer at target objects in the consecutive condition than in the interleaved condition during these learning instances.

Comparisons with Simulations

Table 2 displays the simulated performances of the learning strategies under consideration and the performance of adults participants. Levene's test indicated that the assumption of homogeneity of variance had been violated for these data, $F(2,243) = 8.032, p < .001$. Therefore, we used Welch's adjusted F-ratio provided in the one-way ANOVA output to test for differences between groups. There was a statistically significant main effect of the type of strategy used on word learning performance with consecutive exposures, $F(2,151) = 268.44, p < .001, \omega^2 = .68$. Planned contrasts revealed that participants performed significantly better in the consecutive condition of the word learning procedure than the

Table 2: Learning performance of the one-exposure learner (Smith et al., 2009), propose-but-verify learner (Trueswell et al., 2013), adult participants, and perfect XSL (Smith et al., 2011).

	Consecutive		Interleaved	
	M	*SE*	*M*	*SE*
One-exposure	0.93	.056	0.93	.056
Prose-but-verify	2.11	.097	2.11	.097
Participants	3.40	.093	2.67	.122
Perfect XSL	4.00	-	4.00	-

Note. $N = 82$ for participants and simulations.

one-exposure learner, $t(133) = 22.73, p < .001, r^2 = .79$ and the propose-but-verify learner, $t(161) = 9.63, p < .001, r^2 = .36$. A one-sample t-test showed that the participants performed significantly worse than a perfect cross-situational learner, $t(81) = -6.41, p < .001, d = 1.42$.

When the learning performance on the interleaved condition was entered separately in the model, Levene's test showed that the assumption of homogeneity of variances had been violated, $F(2,243) = 28.024, p < .001$. Welch's test results showed a statistically significant main effect of the type of strategy used on word learning performance with interleaved exposures, $F(2,154) = 220.63, p < .001, \omega^2 = .48$. Planned contrasts revealed that participants performed significantly better in the interleaved condition of the word learning procedure than the one-exposure learner, $t(138) = 13.02, p < .001, r^2 = .76$ and the propose-but-verify learner, $t(161) = 13.02, p < .001, r^2 = .29$. A one-sample t-test showed that the participants performed significantly worse than a perfect cross-situational learner, $t(81) = -10.97, p < .001, d = 2.44$.

Discussion

In this paper, we explore cross-situational word learning under different cognitive loads without using a forced-choice paradigm. To this aim, we designed an experiment in which we tracked participants' eye movements – a method that has proven adequate for studying this kind of learning behavior (Yu & Smith, 2011). To obtain insights into the strategies people use for word learning, we compared the experimental results with simulations in which various learning strategies were modeled.

In order to prove that our participants were collecting information from multiple exposures, we needed to show that they outperform the one-exposure learning strategy proposed by Smith et al. (2009). We confirm that our participants significantly outperformed the one-exposure learner, suggesting that they integrated information about potential word-referent mappings cross-situationally. This outcome is in line with Smith et al. (2009). In order to show that our participants were tracking multiple hypotheses simultaneously, we needed to demonstrate that they outperform the propose-but-verify learning strategy by Trueswell et al. (2013). Our participants indeed learned significantly more words than the propose-but-verify simulation. This finding is inconsistent with Trueswell et al. (2013), who found that their participants learned target words using a propose-but-verify strategy in which they tracked only one hypothesis at a time. The fact that our participants outperformed the propose-but-verify learner indicates that participants integrated more information across situations than just information about one possible word-meaning mapping at a time. This result corresponds with studies which show that adult learners do indeed infer and track multiple hypotheses across situations (Gillette et al., 1999; Piccin & Waxman, 2007; Yu & Smith, 2007; Smith et al., 2009, 2011). Consistent with previous research, our participants did not reach the performance level of a perfect XSL account (Smith et al., 2011), suggesting that information loss occurs across exposures. The fact that learning instances were sometimes presented consecutively or interleaved with learning instances for other words did not prevent participants from using XSL. Contrasting with what Trueswell et al. (2013) suggest, XSL seems to be at play even when participants are presented with a limited number of learning instances that are interleaved with learning instances for other words. However, participants learned slower and less effective in the interleaved condition than in the consecutive condition, presumably due to memory constraints.

Whereas Trueswell et al. (2013) and Smith et al. (2011) used a forced-choice paradigm (i.e. participants were told to click on one of the referents on each exposure), learners usually also have the option not to select a referent on first exposures to target words. Participants were required to propose a potential word-meaning mapping in situations where they would not naturally select a referent due to high uncertainty. We believe that this paradigm might have biased the learning behavior of Trueswell et al.'s participants, and consequently may have induced a propose-but-verify strategy. In our look-and-listen paradigm, the preference for a referent, or multiple referents, was measured indirectly through gaze behavior, relieving participants of making a decision on first exposures to target words. The findings demonstrate that -on average- our participants reached a performance that suggests they track multiple hypotheses simultaneously.

Koehne et al. (2013) reported that their participants memorized more than one potential meaning for a target word, unlike a strict propose-but-verify theory would predict. As participants mostly memorized objects they had proposed during training, they interpret their findings in the light of a multiple-proposal account for word learning. However, in half of the conditions participants selected objects that occurred frequently from early on in the experiment above chance-level on test trials, which we believe is consistent with a XSL account for word learning. In order to distinct between (weakened forms of) XSL and propose-but-verify, a deeper analysis of gaze behavior during the word learning procedure is required. Preliminary results of such an analysis suggest that the majority of our participants used a variety of XSL strategies (cf. Smith et al., 2011), but did not rule out the possibility that some participants used a propose-but-verify strategy.

To conclude, our experiment demonstrates that adults track multiple hypotheses simultaneously when they are not forced to choose referents during word learning. They do so even under cognitively demanding circumstances, but they learn faster and more effectively when learning instances for target words are displayed consecutively than when they are interleaved with learning instances for other words.

Acknowledgments

Our gratitude goes to Kenny Smith, Andrew Smith, and Richard Blythe for making their stimuli set available to us and to Rein Cozijn for his invaluable advice on this study.

References

Cozijn, R. (2006). Het gebruik van oogbewegingen in leesonderzoek. *Tijdschrift voor taalbeheersing*, *28*(3), 220-232.

Gillette, J., Gleitman, L., Gleitman, H., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135-176.

Koehne, J., Trueswell, J., & Gleitman, L. (2013). Multiple Proposal Memory in Observational Word Learning. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (p. 805-810). Austin, TX: Cognitive Science Society.

Markman, E., & Wachtel, G. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121-157.

Piccin, T., & Waxman, S. (2007). Why nouns trump verbs in word learning: New evidence from children and adults in the human simulation paradigm. *Language Learning and Development*, *3*(4), 295-323.

Quine, W. (1960). *Word and object.* The MIT Press.

Siskind, J. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1-2), 39-1.

Smith, K., Smith, A., & Blythe, R. (2009). Reconsidering human cross-situational learning capacities: A revision to Yu & Smith's (2007) experimental paradigm. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (p. 2711-2716). Austin, TX: Cognitive Science Society.

Smith, K., Smith, A., & Blythe, R. (2011). Cross-situational learning: an experimental study of word-learning mechanisms. *Cognitive Science*, *35*(3), 480-498.

Trueswell, J., Medina, T., Hafri, A., & Gleitman, L. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126-156.

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

Yu, C., & Smith, L. (2011). What you learn is what you see: using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*(2), 165-180.