Sentence Processing Mechanisms Influence Cross-Situational Word Learning

Judith Köhne & Matthew W. Crocker
Department of Computational Linguistics
Saarland University
Saarbrücken, Germany
{judith, crocker}@coli.uni-saarland.de

Abstract

Word learning has traditionally examined separately the role of constraints provided by the visual context (e.g. *cross-situational learning*) and the linguistic context (e.g. *syntactic bootstrapping*). We suggest that the combined investigation of these learning scenarios is important: Firstly, to determine whether cross-situational word learning applies when words are presented in sentences and, secondly, to illuminate possible interactions of linguistic and situational learning mechanisms. We conducted three experiments to examine the role of visual and linguistic contextual constraints during foreign language word learning. In particular, our studies show that, given a visual context, syntactic verb-argument constraints together with knowledge about plausible real-world action-object relations help to further enhance cross-situational word learning.

Keywords: Cross-situational word learning in context; sentence processing; verb-derived expectations;

Introduction

Adults' foreign language learning often happens in a planned and incremental way to systematically gain increasing command of the language's structures. Parts of the vocabulary are learned very explicitly via vocabulary lists. When it comes to using and improving a foreign language in a natural situation, within an actual speech community, however, the language novice faces a less controllable, more diverse situation. When trying to understand and learn new words, there are two challenges: Firstly, words are often embedded in complex linguistic contexts and, secondly, there is a rich but ambiguous visual context containing possible world referents (*referential uncertainty*, Gleitman, 1990). One important mechanism for dealing with referential uncertainty is to keep track of words and referents co-occurring over different contexts. As previous research shows, adults as well as children are able to exploit such cross-situational learning analyses (*cross-situational word learning*, CSWL, e.g., Quine, 1960; Yu and Smith, 2007; Vouloumanos and Werker, 2009). In a study by Yu and Smith (2007), participants were asked to learn novel names for novel objects. Within a single trial, participants were exposed to 2-4 auditorily presented nouns, unconnected to each other, and the equal number of visually presented objects. Despite the referential uncertainty in each trial, participants were able to learn noun-object mappings by exploiting cross-trial co-occurrences.

In most CWSL studies, words are not presented as parts of sentences. This idealization has drawbacks, however: Firstly, language is not presented in its natural complexity (i.e., with sententially embedded words) and, secondly, possibly useful constraints provided by the linguistic context are intentionally withheld. This means that learning tasks are potentially oversimplified in some respects and overcomplicated in others. There is some evidence that adults are able to make use of the linguistic context that words come together with to understand the language input, instead of being distracted by it. Lee and Naigles (2008), for instance, show that the verb frame of a sentence helps verb learning (*syntactic bootstrapping*, Landau and Gleitman, 1985; Fisher, 2002; Lidz, Gleitman, and Gleitman, 2003). These kinds of studies, however, are usually not visually situated.

Some sentence processing mechanisms that are generally automatically applied by adult native speakers may also interact when dealing with foreign language input. As Altman and Kamide (1999) have shown, for instance, native speakers rapidly make inferences about linguistically upcoming referents in a sentence, given a restrictive verb (such as *eat*) and a visual scene. We investigate the hypothesis that learners may use similar on-line mechanisms when learning novel nouns. Specifically, we investigate whether such on-line predictions influence CSWL by reducing the size of sets of potential world referents a novel noun refers to.

In this paper, three adult language-learning eye-tracking experiments using a pseudo-natural language (modified Indonesian) are presented, addressing three central hypotheses: 1) CSWL mechanisms operate successfully when novel nouns are embedded in sentences. 2) Verb-driven, anticipatory expectations based on semantic verbal restrictions guide learners' (visual) attention. 3) Verb-driven, anticipatory expectations based on semantic verbal restrictions identify subsets of world-referents that novel nouns are likely to denote, thereby constraining CSWL.

Experiment 1

We investigated these issues with a stepwise learning procedure. Participants first learned restrictive verbs and were then exposed to novel nouns, embedded in spoken subject-verb-object (SVO) sentences as syntactic subjects (referring to characters) and syntactic objects (referring to objects), and depicted on scenes.

Methods

Participants 32 German native speakers took part in the experiment, 8 of which had to be excluded due to technical problems. Data of 24 participants was analyzed (17 female).

Design, Materials & Procedure The language consisted of six restrictive verbs (three food verbs like *bermamema*, 'eat', and three clothing verbs like *melimema*, 'iron'), twelve nouns

(six referring to human characters such as *badut*, 'clown', three to food objects such as *sonis*, 'sausage', and three to clothing items, e.g. *oblung*, 't-shirt') and one article that preceded all nouns (*si*, 'the'). The language was based on Indonesian (word order, article, parts of the words).

The experiment comprised three phases: isolated verb learning, situated noun learning, and vocabulary testing. In Phase 1, participants learned verbs by being exposed to static depictions of actions presented together with the corresponding spoken verb. Each action was named ten times. Then knowledge of verbs was tested: Participants were presented a picture not seen before and asked to pronounce the matching verb. Feedback was provided. The eye-tracker was adjusted and verbs together with depictions were presented again.

Phase 2 consisted of the sentence-comprehension and noun-learning phase: Semi-natural scenes and spoken sentences were presented (sentence start 1s after picture). Scenes depicted the target character and the target object (the named referents), as well as one distractor character and one distractor object, together with background. There was always one food item and one clothing item. Sentences were constructed using the already learned verbs and novel nouns. Word order was SVO. The syntactic subject denoted the target character and the syntactic object referred to the target object, either the food or the clothing one, corresponding to verb type (see example in Figure 1). People were not explicitly told the word order. There were 36 trials (randomized in order), each object and each character was named six times (and each one was shown twelve times). Participants were asked to understand the sentences and learn the unknown words. Eye-movements were measured.

In Phase 3, a forced-choice vocabulary test with 12 trials, one for each new noun, was performed. Pictures for the forced-choice vocabulary test showed 4 potential referents (= characters and objects) and were presented together with a spoken noun. Combinations of the four options differed but there was always at least one competitor of the same kind (character, food, or clothing item, respectively). Participants had to mouse-click onto the appropriate picture. Learning performance was the main measurement of interest for Phase 3. The experiment lasted about 30 minutes.

Predictions Hypothesizing that participants understand the SVO-sentence structure and have similar gaze behavior as native comprehenders, we expected more looks to characters than objects during NP1 and more looks to objects than characters during NP2. Secondly, we hypothesized that to identify character referents and learn their names, participants would exploit cross-situational analysis (Hypothesis 1). This predicts differences between looks to target and distractor characters to emerge over time during NP1: While in the very beginning participants have no hint which character NP1 referred to, tracking co-occurences of character names and depictions over trials makes it possible to identify the target. This increase in looks to the target should also become visible in the averaged data. We further hypothesized that verb

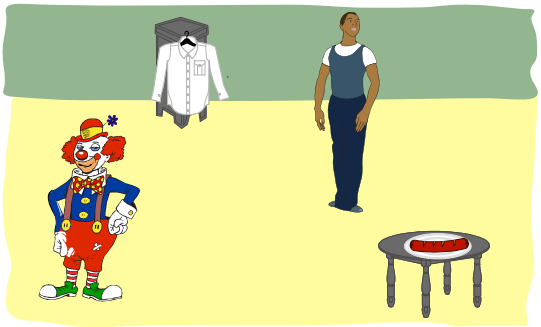

Figure 1: Example Item Experiment 1
Si badut bermamema si worel.
'The clown will eat the sausage.'

restrictions would be exploited quickly to identify object referents and learn their names, possibly additionally to CSWL (Hypotheses 2 & 3). That means that during the verb and NP2, targets should be inspected much more than distractors even early in Phase 2. Our hypothesis that verb restrictions provide additional cues regarding target objects (Hypothesis 3) moreover predicts that object names are learned better than character names overall.

Data Analysis, Results, & Discussion

The vocabulary test revealed a noun-learning rate that is well above chance (about 55% with a baseline of 25%, $t = 9.28, p < .001$). When analyzing only the data of the participants who actually learned all verbs in Phase 1, $N = 15$, it was 64% ($t = 8.59, p < .001$). Numerically, object names were learned better than character names but this difference did not reach significance (all: $t = .90, p = .38$; only good verb learners: $t = 1.38, p = .191$).

For eye-movement analysis we examined trials with at least one inspection on our regions of interest (ROI; target character, distractor character, target object, distractor object) for three time periods linked to the unfolding sentence (from onset of NP1 to onset of verb (V), from onset of V to onset of NP2, and from onset of NP2 to offset of NP2). All time periods were shifted such that they started 200ms later than the actual starting points in the speech stream because planning of saccades takes people about that much time. We conducted logistic regressions by entering the binomial data (fixation or no fixation at certain time to a specific ROI) into linear mixed effect models with logit link function (from the lme4 package in R, Bates, 2005). Participant and item were considered as random factors. To see whether the fixed factor (ROIs) had a main effect (i.e. whether including the factor significantly improved the predictive power of the model, regarding where people looked) we compared between the models that include and exclude this factor with a Chi-Square test (Baayen, Davidson, & Bates, 2008). Contrasts between levels of a factor (i.e. single ROIs) were investigated by studying the ratio of regression coefficients and standard errors since the p-values produced by lmers (Wald z test) are anti-conservative

(Baayen et al., 2008): If the coefficient is greater than the standard error times two, the comparison is considered to be reliable. Tables of these statistical comparisons are given below. The formulas describing the lmer models are of the following form: dependent variable (inspections during time periods) is a function of ($\sim$) the independent variable (ROI) plus random effects (subjects and items).

Table 1: Lmer models for inspections on characters vs. objects (m1) and targets vs. distractors (m2) during time periods (Experiment 1)
m1/m2: $InspectionsduringNP1/V/NP2 \sim 1 + ROI + (1|sub) + (1|item), family = binomial(link = "logit")$

		Predictor	Coef.	SE	Wald z	p
m1						
1	*NP1*	(Int) (char)	1.953	0.150	13.019	$< .001$
2		objects	-0.859	0.129	-6.647	$< .001$
3	*V*	(Int) (char)	0.224	0.116	1.928	$< .100$
4		objects	1.317	0.114	11.548	$< .001$
5	*NP2*	(Int) (char)	-0.253	0.146	-1.732	$< .100$
6		objects	0.492	0.101	4.872	$< .001$
m2						
7	*NP1*	(Int) (targ)	0.925	0.145	6.378	$< .001$
8		distractor	-0.112	0.111	-1.006	$= .315$
9	*V*	(Int) (targ)	0.593	0.122	4.859	$< .001$
10		distractor	-0.382	0.102	-3.734	$< .001$
11	*NP2*	(Int) (targ)	-0.557	0.107	-5.195	$< .100$
12		distractor	-0.257	0.105	-2.452	$< .050$

Eye-movements suggest that participants quickly understood the sentence structure: There were reliably more inspections on the characters than inspections on the objects during NP1 (Table 1, rows 1-2) and reliably more inspections on the objects than on the characters in the V interval (rows 3-4) and NP2 (rows 5-6). Moreover, the target character was inspected more often than the distractor character in NP1 (rows 7-8) and the target object was looked at more than the distractor object during NP2 (rows 11-12). This supports the hypothesis that participants succeeded in identifying the targets over the experiment. Furthermore, during V, the target object was also looked at reliably more than the distractor (rows 9-10). This likely reflects an anticipatory effect based on semantic verb restrictions. Also, the difference between looks to target and distractor objects during NP2 was greater than the difference between target and distractor characters during NP2, suggesting that verb restrictions contributed to an improved identification of objects during on-line processing (see timegraph in Figure 2).

Summarising, we found evidence that adults can learn nouns cross-situationally when words are embedded in sentences and referents are embedded in scenes, and further that they rapidly exploit semantic verb restrictions to identify referents on-line. We replicated these results with even better learning rates ($72\%, t = 8.249, p < .001$) and for another word order (OVS, with a learning rate of 51%, $t = 3.840, p < .001$) in a follow-up experiment. It is unclear, however, whether the on-line predicting of the referent has an effect on noun learning. We take this up in Experiment 2.

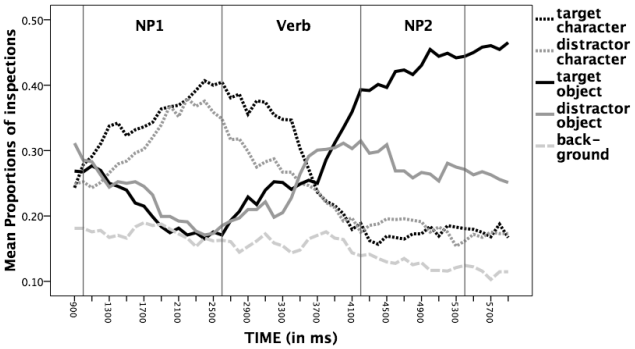

Figure 2: Timegraph Experiment 1

Experiment 2

In Experiment 2 we manipulated the degree of verb restriction to study the interaction of CSWL and verb-derived inference learning (Hypothesis 3). The focus therefore was on object learning rather than character learning.

Methods

Participants 50 German native speakers took part in the experiment (18 excluded due to bad verb learning and technical problems). Data of 32 participants was analyzed (23 female).

Materials & Procedure The language consisted of six verbs, 14 nouns, and the same article as in the other experiments. There were 2 non-restrictive verbs (e.g., *take*) and 4 restrictive verbs: either 2 food verbs (e.g., *eat*) and 2 clothing verbs (e.g., *iron*) or the 2 food verbs and 2 container verbs (e.g., *fill*), depending on list. The nouns denoted 2 characters (man and woman) and 12 objects: 4 food items (e.g., *broccoli*), 4 clothing items (e.g., *trousers*), and 4 container items (e.g., *vase*). Word order was SVO.

The experiment consisted of five parts with very similar procedures as in Experiment 1. The main difference was that instead of one sentence comprehension phase and one vocabulary testing part, there were two each (Blocks 1 and 2). The whole experimental sequence comprised: verb learning and testing, eye-tracker preparation, and verb repetition (Phase 1); sentence comprehension (noun learning) Block 1 (Phase 2); vocabulary test Block 1 (and verb repetition) (Phase 3); sentence comprehension (noun learning) Block 2 (Phase 4); vocabulary test Block 2 (Phase 5).

Phase 1 resembled Phase 1 in Experiment 1, except that we used animated verb-learning and verb-testing pictures to improve recognizability of the actions. In Phase 2 and 4, items were manipulated according to one three-level within-participant factor (Degree of referential uncertainty). There were three conditions: the no-referential-uncertainty condition (Condition 1), the low-referential-uncertainty condition

(Condition 2), and the high-referential-uncertainty condition (Condition 3). Firstly, the conditions differed with regard to verb type: In Condition 1 and 2, a restrictive verb was used, in Condition 3, it was a non-restrictive verb. Secondly, there were differences in the visual scenes. Images always depicted one character and four objects embedded in a simple indoor scene. One of the objects was the target object. The others were competitors (= potential world referent except the target) and distractors (= objects which are not potential world referents). The combination of competitors and distractors depended on the condition the item was in: In Condition 1, there was no competitor since the verb was restrictive (e.g. *eat*) and only the depicted target fulfilled verb constraints (e.g. *food*). In Condition 2, there was one competitor: The verb was restrictive but there was one depicted object in addition to the target which was a member of the required semantic class. In Condition 3, there were three competitors because the verb was non-restrictive and did not semantically constrain the category of potential referents denoted by the post-verbal argument (see Table 2). In Block 1 (Phase 2), no target was a competitor in another trial to make sure participants could not exclude competitors based on other learned words. In Block 2 (Phase 4), however, learning was potentially simplified via the possibility to exclude already learned mappings. The 48 trials (24 per Block) were presented randomized in order with each noun repeated four times. Participants were told that sentences were of the form 'someone VERBs something'. We monitored eye-movements in Phases 2 and 4.

Table 2: Conditions Experiment 2
(Number of potential referents on scene (Column 4) as a result of verb type (Column 2) and number of competitors (Column 3))

Condition	Verb	Competitors	= Pot. referents
1: No-ref. unc.	restr.	0	1
2: Low-ref. unc.	restr.	1	2
3: High-ref. unc.	non-r.	3	4

For the forced-choice vocabulary tests (Phases 3 and 5) there were six depictions presented on the screen: the target (*e.g. tomato*) and another instance of the targets category (e.g. *broccoli*), two objects of one of the two other categories (e.g., *shirt* and *skirt*), and two characters. Additionally to the mouse clicks, we introduced a confidence rating to have another, more sensitive measurement because there were only two nouns to be learned per condition: Participants were encouraged to press a number on the keyboard (between 1-9) to indicate how sure they were about their choice of a referent.

Predictions We hypothesized that selectional verb restrictions help identifying target referents and interact with CSWL (Hypothesis 3). In particular, we hypothesized, firstly, that during Phases 2 and 4, verb restrictions narrow down the search space in Condition 1 (from four to one since there was no competitor) and Condition 2 (from four to two since there was one competitor) (see Table 2); Secondly, that participants additionally conduct CSWL in Condition 2; And thirdly, that participants conduct only CSWL in Condition 3. Our predictions were, therefore, that noun learning rates and confidence ratings, as reflected in the vocabulary tests, would be highest for objects in Condition 1 and lowest for objects in Condition 3. With regard to eye-movements in Phases 2 and 4, our hypotheses predict differences for conditions during NP2: a clear preference for inspecting the target in Condition 1, a preference to inspect the target but a secondary preference to inspect the competitor in Condition 2, as well as a less strong preference for target inspection and an equally strong consideration of all competitors in Condition 3.

Finally, we hypothesized that in Block 2 participants can exclude those objects as potential referents, which have been already linked to a world-word-mapping in Block 1 (assuming the use of the principle of mutual exclusivity, Markman and Wachtel, 1988). This predicts an enhanced noun learning in Block 2 compared to Block 1.

Data Analysis, Results, & Discussion

Noun learning was reliably better than chance (25%) for all groups of interest and correlated positively with confidence ratings ($r = .452, p < .001$, see Table 3).

Learning was clearly better in Block 2 than Block 1 (all conditions: $\chi(1) = 30.77, p < .001$; Condition 1: $\chi(1) = 6.31, p < .05$; Condition 2: $\chi(1) = 10.17, p < .01$; Condition 3: $\chi(1) = 16.57, p < .001$). The same was true for confidence ratings (all conditions: $\chi(1) = 12.85, p < .001$; Condition 1: $\chi(1) = 5.42, p < .05$; Condition 2: $\chi(1) = 5.48, p < .05$; Condition 3: $\chi(1) = 10.69, p < .01$).

The direction of the differences between noun learning success and confidence ratings in the three conditions was as expected: Nouns were learned best and the decisions were rated highest in Condition 1 and worst in Condition 3. This was true for both blocks together as well as for Blocks 1 and 2 separately. We analyzed both values with linear mixed effect models, using logistic regression for the categorical learning rates (logit link function) and linear regression for the continuous confidence ratings, with participant and item as random factors. For confidence ratings we calculated Monte Carlo Marcov Chain values (MCMCs) whose p-values are a good estimate of the factor's significance (but are only applicable for continuous variables), (Baayen et al., 2008). Analyses did not reveal significant main effects for noun learning rates but did for confidence ratings. There were reliable differences in confidence ratings between single conditions: Condition 1 and Conditions 3 in all parts (Block 1, 2, and 1+2), Condition 1 and Condition in 2 in 1+2 and marginally in Block 2, and between Conditions 2 and 3 in all parts (Table 4: numbers in both blocks taken together).

Eye-movements were analyzed as in Experiment 1. Considering all conditions, the eye-gaze pattern for all parts of the

Table 3: Noun learning percentages (t-tests against chance 25%) / confidence ratings, Experiment 2

	Blocks 1+2	Block 1	Block 2
all	$72\%(t(62) = 12.18, p < .001)/5.73$	$62\%(t(62) = 6.90, p < .001)/5.06$	$83\%(t(62) = 14.24, p < .001)/6.39$
Cond1	$77\%(t(62) = 10.04, p < .001)/6.98$	$69\%(t(62) = 6.24, p < .001)/6.34$	$85\%(t(62) = 12.56, p < .001)/7.5$
Cond2	$74\%(t(62) = 9.25, p < .001)/6.42$	$64\%(t(62) = 5.19, p < .001)/5.88$	$85\%(t(62) = 11.34, p < .001)/6.8$
Cond3	$66\%(t(62) = 7.43, p < .001)/5.4$	$52\%(t(62) = 3.49, p < .001)/4.45$	$80\%(t(62) = 8.69, p < .001)/6.02$

Table 4: Lmer models & p-Values from MCMC sampling for confidence ratings, conditions 1-3 (Exp 2, both blocks)
m1: $condition \sim 1 + confidencerating + (1|sub) + (1|item)$

	Predictor	Coefficient	SE	t	MCMCmean	pMCMC	$Pr(>\|t\|)$
confidence ratings	(Intercept) (Condition1)	7.0575	0.3804	18.554	7.0303	$< .001$	0.0000
	Condition2	-0.6371	0.2869	-2.221	-0.58376	$< .100$	0.0272
	Condition3	-1.7530	0.30235	-5.799	-1.7027	$< .001$	0.0000

Table 5: Lmer models for inspections on target vs. distractors (m1) and distractor1/competitor vs. rest (m2) during NP2, conditions 1-3 (Exp 2, both blocks together)
m1/m2: $InspectionsduringNP2 \sim 1 + ROI + (1|sub) + (1|item), family = binomial(link = "logit")$

		Predictor	Coef.	SE	Wald z	p
m1						
1	*Cond1*	(Int) (tar)	-0.314	0.130	-2.406	$< .050$
2		char	-1.310	0.174	-7.540	$< .001$
3		di1	-0.685	0.171	-4.003	$< .001$
4		di2	-0.741	0.172	-4.306	$< .001$
5		di3	-0.506	0.174	-2.910	$< .010$
6	*Cond2*	(Int) (tar)	-0.124	0.121	-1.032	$= .300$
7		char	-1.426	0.174	-8.210	$< .001$
8		di1	-0.509	0.162	-3.134	$< .010$
9		di2	-1.325	0.186	-7.134	$< .001$
10		di3	-1.260	0.190	-6.642	$< .001$
11	*Cond3*	(Int) (tar)	-0.239	0.129	-1.859	$< .100$
12		char	-1.562	0.179	-8.709	$< .001$
13		di1	-0.172	0.156	-1.102	$= .270$
14		di2	-0.505	0.165	-3.058	$< .010$
15		di3	-0.983	0.178	-5.533	$< .001$
m2						
16	*Cond1*	(Int) (di1)	-0.998	0.141	-7.090	$< .001$
17		tar	-0.691	0.171	4.038	$< .001$
18		char	-0.629	0.181	-3.469	$< .001$
19		di2	-0.501	0.180	-0.317	$= .751$
20		di3	0.178	0.182	0.983	$= .326$
21	*Cond2*	(Int) (com)	-0.633	0.131	-4.840	$< .001$
22		tar	0.509	0.162	-8.208	$< .010$
23		char	-0.917	0.181	-5.070	$< .001$
24		di2	-0.816	0.192	-4.249	$< .001$
25		di3	-0.751	0.196	-3.826	$< .001$
26	*Cond3*	(Int) (di1)	-0.411	0.130	-3.171	$< .010$
27		tar	0.172	0.156	1.102	$= .270$
28		char	-1.389	0.180	-7.715	$< .001$
29		di2	-0.333	0.166	-2.001	$< .050$
30		di3	-0.811	0.179	-4.524	$< .001$

experiment resembles that of Experiment 1: We found referential inspections of the character during NP1, verb-driven anticipation of the target object(s) in verb region, and referential inspections of the target object in NP2.

More interestingly, differences between conditions for (referential) inspections in NP2 support the offline results: In Condition 1, the target was inspected reliably more than the character and the distractors (in all parts of the experiment) (Table 5, rows 1-5, for Blocks 1 and 2 together). In Condition 2, the target was inspected most, too (rows 6-10); However, the competitor was also inspected reliably more than the character and the distractors. The difference between looks to target and competitor was not significant in Block 1 but was in Block 2 and both blocks taken together (rows 21-25). For Condition 3, the target was inspected reliably more than the character or the distractors as well, except that the difference between looks to the target and to one distractor was significant only in Block 2, but neither in Block 1 nor in both blocks taken together (rows 11-15). This distractor shared category with the target. There were also significantly more looks to this distractor than to the other distractors and the character in Block 2 (but not for both blocks, see rows 26-30). The gaze pattern for Condition 3 is somewhat unexpected but interesting as it suggests that participants learned a new co-occurrence restriction for verbs in Block 2 (e.g., container objects and *take*) - although the verbs were non-restrictive, the distractor of the target category was preferred over the other distractors (which were of categories associated with other, restrictive, verbs).

The second experiment revealed clear effects of condition in on-line and off-line data showing that, firstly, referents are identified better when verbs provide information about the referent's category (better learning rates and confidence ratings in Condition 1 and 2 than in Condition 3) and, secondly, that cross-situational word learning interacts with the exploration of verb restrictions in that verb restrictions narrow

down the search space, lowering referential uncertainty (better learning rates and confidence ratings for Condition 2 than 3). As in Experiment 1, trials in Condition 1 made clear that verb restrictions can narrow down the number of potential referents to one, which means that there is a situation close to fast-mapping. Eye-movements during NP2 support the results except that there was an unexpected preference to look at both members of target category in Condition 3. We attribute this to spontaneous verb-argument category learning.

Summary & General Discussion

Two foreign-language learning experiments with an incremental learning scenario were conducted in order to study the influence of semantic verb restrictions on identifying world referents and learning world-word mappings. In Experiment 1, we found that nouns which are sententially embedded are successfully learned cross-situationally (with SVO and OVS sentences) and that participants additionally exploited verb restrictions rapidly to identify post-verbal referents. In Experiment 2, we additionally found evidence for the claim that verb restrictions interact with and improve Cross-Situational Word Learning.

With this investigation we have presented evidence for the claim that adult sentence processing mechanisms interact with statistical word learning and that foreign language word learners can benefit from exploiting linguistic and visual contextual cues. In particular, we revealed that semantic verb restrictions together with knowledge about plausible arguments reduces the set of potential (visual) referents, thus simplifying CSWL complexity. This highlights the cooperation of multiple learning mechanisms in situated word learning. Our findings are consistent with recent word learning models which combine co-occurrences frequency analysis with other, in particular situational and knowledge-based cues (Frank, Goodman, & Tenenbaum, 2009; Yu & Ballard, 2007).

Acknowledgments

The research reported of in this paper was supported by IRTG 715 "Language Technology and Cognitive Systems" funded by the German Research Foundation (DFG).

References

Altman, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Bates, D. (2005). Fitting linear mixed models in r. *R News*, *5*, 27–30.

Fisher, C. (2002). Structure limits on verb mapping: The role of abstract structure in 2.5-year- olds interpretations of novel verbs. *Developmental Science*, *5*, 55–64.

Frank, M., Goodman, N., & Tenenbaum, J. (2009). Using speaker's referential intentions to model early cross-situational word learning. *Psychological Science*, *20*, 578–585.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55.

Landau, B., & Gleitman, L. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.

Lee, J., & Naigles, L. (2008). Mandarin learners use syntactic bootstrapping in verb acquisition. *Cognition*, *106*, 1028–1037.

Lidz, J., Gleitman, H., & Gleitman, L. (2003). Understanding how input matters: Verb-learning and the footprint of universal grammar. *Cognition*, *87*, 151–178.

Markman, E., & Wachtel, G. (1988). Childrens use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*, 121–157.

Quine, W. (1960). *Word and object*. Cambridge, MA.

Vouloumanos, A., & Werker, J. (2009). Infant's learning of novel words in a stochastic environment. *Developmental Psychology*, *45*, 1611–1617.

Yu, C., & Ballard, D. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*, 2149-2165.

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