The interplay of cross-situational word learning and sentence-level constraints

Judith Koehne[a] and Matthew W. Crocker[b]

[a] Institute for Research in Cognitive Science, University of Pennsylvania *

[b] Department of Computational Linguistics, Saarland University

Corresponding author:

Judith Koehne

+49 − 1777879436

judithkoehne@gmail.com

SUGGESTED RUNNING HEAD: Statistical word learning and sentential cues

Keywords: Interaction of word learning mechanisms; Statistical word learning; Sentential cues; Visual world paradigm

*now: Department of Psychology, University of Bamberg

Abstract

A variety of mechanisms contribute to word learning. Learners can track co-occurring words and referents across situations in a bottom-up manner (*cross-situational word learning*, CSWL). Equally, they can exploit sentential contexts, relying on top-down information such as verb-argument relations and world knowledge, offering immediate constraints on meaning (*word learning based on sentence-level constraints*, SLCL). When combined, CSWL and SLCL potentially modulate each other's influence, revealing how word learners deal with multiple mechanisms simultaneously: Do they use all mechanisms? Prefer one? Is their strategy context dependent? Three experiments conducted with adult learners reveal that learners prioritize SLCL over CSWL. CSWL is applied in addition to SLCL only if SLCL is not perfectly disambiguating, thereby complementing or competing with it. These studies demonstrate the importance of investigating word-learning mechanisms simultaneously, revealing important characteristics of their interaction in more naturalistic learning environments.

1 Word learning in adults

While word learning is a difficult task, language novices come equipped with a range of devices to meet this challenge: Firstly, they can perceive and integrate a range of available information sources such as the linguistic context and the visual environment. Secondly, they can combine such situational cues with their prior experience, with both language and the world, to constrain word meanings. One way language novices have been shown to apply these skills is to observe the visual world to identify the object or event a novel word may refer to (*observational word learning*, Carey, 1978). To reduce the vast ambiguity regarding potential mappings between spoken words and visual referents (*world-word mappings*), learners can track the frequencies of their co-occurrences across situations in a bottom-up fashion (*cross-situational word learning*, CSWL, e.g., Quine, 1960; Yu & Smith, 2007). CSWL is a potentially powerful word learning mechanism which has received considerable attention in recent years. When encountering a novel word embedded in a sentence, learners can moreover use the sentential contexts in a top-down manner to make inferences about its meaning. Such inferences may be based on the learner's prior knowledge about sentence level constraints, plausible states and relations in the world, or both (e.g. Field, 2004; Landau & Gleitman, 1985). We will henceforth refer to this kind of word learning as *sentence-level constraint learning* (*SLCL*). SLCL has convincingly been shown to play an important role, in particular for verb learning (*syntactic bootstrapping*, Landau & Gleitman, 1985; e.g., Fisher, 2002). The present investigation seeks to extend these previous results to the learning of nouns based on the selectional restrictions of verbs.

CSWL and SLCL are particularly interesting exemplars of two word learning mechanisms which are fundamentally different in nature: While CSWL operates bottom-up and probabilistically to help learners to identify likely word-world mappings over time and across situations, SLCL provides top-down semantic constraints on word meaning, facilitating a more deterministic and thus immediate identification of a novel word's referent possible.

While many studies support the importance of mechanisms such as CSWL and SLCL for word learning, little is known about the interplay of these and other word learning mechanisms. In contrast to early stages of child word learning, which are likely to rely on observational (and social) cues only (because information about syntactic relations is not yet available; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005), relying on only one mechanism seems implausible for first or second language word learning in adults or older children. Thus, it is crucial to examine the simultaneous deployment of relevant learning mechanisms in order to attain a precise understanding of how words are learned in real-world word-learning situations. Studying learning mechanisms in isolation potentially makes unprecise assumptions on three levels: Firstly, *the importance of each learning mechanism* is potentially wrongly estimated because different mechanisms may influence each other by either reducing or enhancing one another's effect, depending on the learning scenario. Secondly, isolating learning cues also means that *the learning environment* is neither presented in its natural complexity (e.g., isolated words rather than natural sentences, in CSWL studies) nor in its natural richness (e.g., the absence of useful constraints provided by the sentential context). This means that learning tasks are potentially oversimplified in some respects and overcomplicated in others. Thirdly, *the language learners' strategy* may not be correctly taken into account: On the one hand, giving them the opportunity to make use of all information sources simultaneously may improve their performance, while on the other hand, presenting them such a complex situation may also cause them to ignore available cues or even result in unsuccessful learning. They may attempt to apply all possible mechanisms whenever possible, prefer to always rely on one, ignore all, or alternate between these strategies in a situation-dependent way.

The objective of the studies presented here is to investigate the interaction of the two established word learning mechanisms CSWL and SLCL in order to evaluate their relative importance and the word learner's strategy regarding the use of multiple mechanisms which are different in nature (bottom-up probabilistic vs. top-down deterministic) in more general. In particular, we address two general research questions. Firstly, under which circumstances and how do CSWL and SLCL modulate each other's operation? Secondly, which general strategy of the adult word learner does this reveal? To address these questions, we implemented the three plausible scenarios of interplay that CSWL and SLCL could be in: SLCL and CSWL as complementary (Experiment 1), redundantly applicable (Experiment 2), and conflicting (Experiment 3). In particular, Experiment 1 evaluates whether CSWL and SLCL can be used in a complementary way to jointly identify word meanings. In Experiment 2, we investigate whether learners still rely on both SLCL and CSWL when the two mechanisms provide redundant information or whether one mechanism blocks the use of the other. The aim of Experiment 3 is to examine whether both SLCL and CSWL compete when they are in conflict or whether either mechanism is prioritized.

1.1 Cross-situational word learning

Both children and adults have been shown to successfully learn words in a cross-situational manner: By keeping track of co-occurring visual referents and spoken words across situations, ambiguity concerning potential referent-word mappings can be resolved over time. Fig. 1 illustrates a possible CSWL setting. A learner of English encounters two situations with two spoken words each and two visual referents each. While the correct world-word mappings in each of these two situations are uncertain (*referential uncertainty*), tracking the co-occurrence frequencies across the two situations can help the learner to disambiguate: In this example, the probability (P) of seeing a heart when the word *heart* is spoken is 1 (i.e., the co-occurrence frequency is 100%) whereas the probability of seeing a cloud when *heart* is spoken is only 0.5.

——————————Insert Figure 1 about here ——————————-

While numerous studies have demonstrated that CSWL can be used successfully in various scenarios and with various learner groups (e.g., Siskind, 1996; Akhtar & Montague, 1999; Vogt & Smith, 2005; Yu & Smith, 2007; Smith & Yu, 2008; Vouloumanos, 2008; Vouloumanos & Werker, 2009; Childers & Paik, 2008; Monaghan & Mattock, 2009; Kachergis, Yu & Shiffrin, 2010b; Breitenstein & Knecht, 2003), the precise nature of CSWL mechanisms has just recently become a matter of discussion (Trueswell, Medina, Hafri, & Gleitman, 2013; Yurovsky, Yu, & Smith, 2013). Studies by Vouloumanos (2008) and Vouloumanos & Werker (2009) support the hypothesis that CSWL is, or at least can be, a probabilistic and parallel word learning mechanism. In Vouloumanos' (2008) study, participants were exposed to learning trials consisting of one isolated animated object depiction and one noun. After these learning trials, they were tested on their noun knowledge in a forced-choice vocabulary test with two choices. Results reveal that participants were sensitive to fine-grained differences in co-occurrence frequencies (0% and 10%, 0% and 20%, 0% and 60%, 0% and 80%, 0% and 100%, 10% and 20%, 10% and 60%, as well as 10% and 80%). While these results support very precise tracking of multiple probabilities concerning world-word mappings, there was no ambiguity within learning trials (because trials consisted of each one referent and word). It is therefore unclear based on this study whether learning is similarly parallel if world-word mappings are more ambiguous.

Evidence against a probabilistic and parallel approach to CSWL comes from Trueswell et al. (2013) (see also Medina, Snedeker, Trueswell, & Gleitman, 2011) who did not find learners to maintain more than one hypothesis about a word meaning at a time. In particular, these experiments reveal that when a learner is asked to choose on object in every trial (e.g., with two referents and one novel noun presented), they show no memory for the unselected object(s) as potential referents in later trials. That means that learners stored only one world-word mapping per situation. It is unclear though whether forcing choices may have motivated this learning behavior.

1.2 Word learning based on sentence-level constraints

There are many ways the linguistic context can be helpful for word learners, due to the numerous systematic semantic and syntactic relations between sentence parts and constraints imposed by one part onto another. A verb's arguments (e.g., subject and direct object), for instance, provide rough information about that verb's semantic category (*syntactic bootstrapping*, Landau & Gleitman, 1985). Relatedly, a study by Lee & Naigles (2008) showed, for example, that 2- and 3-year-old Chinese children use the causative subcategorization frame (someone VERBs someone else) to infer causative meanings of novel verbs.

While the opposite relationship between verbs and their arguments have not yet been investigated with regard to word learning, we know from research in sentence comprehension that verbs' information regarding thematic roles and semantic categories of their arguments is rapidly exploited. A visual-word study by Altmann & Kamide (1999) reveals, for instance, that adults immediately use restrictive verbs (such as to eat) to predict the semantic category of an upcoming post-verbal object noun (e.g., something edible) by consulting their world knowledge; more than that, they anticipate concrete referents in visual scenes (e.g., a cake). These findings have been confirmed by numerous experiments (e.g., Kamide, Altmann, & Haywood, 2003; Knoeferle & Crocker, 2007), also for children (Fernald, Zangl, Luz Portillo, & Marchman, 2008) and second-language processing (Wonnacott, Newport, & Tanenhaus, 2007). Our experiments directly evaluate whether such verb-derived constraints on the semantics of an argument give rise to inferences about the meaning of unknown nouns in direct object position so as to facilitate noun-learning.

1.3 The interaction of word-learning mechanisms

Because studies of word learning typically focus attention on a single mechanism, little is known about the effect of using different word-learning mechanisms simultaneously. However, understanding the interplay of different mechanisms is crucial for a complete understanding of word learning: In more natural learning scenarios, it is likely that different mechanisms and cues may for instance complement each other or conflict with one another, potentially modulating the relevance of an individual mechanism. Moreover, it is impossible to fully understand a word learner's overall behavior and capacity if cues and mechanisms are artificially isolated: While the availability of multiple cues may be an advantage, it may also be cognitively demanding and could, in principle, cause the learner to ignore available information and (cognitive) resources or even result in unsuccessful learning.

Gillette, Gleitman, Gleitman, & Lederer, (1999) (see also Snedeker & Gleitman, 2004) provide a rare but insightful example for how word learning cues in adult learning can interact, finding that combined linguistic context (verb frame and lexical information) and scene information contribute to better verb learning than only one of these cues. It is difficult to disentangle, however, whether these cues (in each particular situation) provided rather complementary or redundant information because the precise interplay was not the focus of investigation in their experimental paradigm. Linguistic cues, in particular linguistic distributional regularities, have interestingly also been shown to facilitate CSWL (Monaghan & Mattock, 2012). In an experiment with English adults, Monaghan and Mattock found that word-object pairing was facilitated when the referring words to be mapped onto visual objects were preceded by non-referring words that were regularly distributed analogous to determiners, compared to a learning scenario with irregularly distributed (or no) determiners. A possible situation of a linguistic word-learning cue conflicting with another word-learning cue, on the contrary, has been examined in a study by Nappa, Wessel, McEldoon, Gleitman, & Trueswell (2009): Results from an experiment with 3-5-year old children reveal that syntactic information (word order of known referent names such as *the elephant*) can override the use of gaze cues (speaker inspecting named referents in the corresponding or opposite order) in verb learning. The authors explain this finding by the high reliability of word-order information in English which is more consistent than speakers' gaze. These results suggest that different word learning mechanisms in different scenarios may interact in very different ways.

It is therefore also an interesting question how the two introduced mechanisms CSWL and SLCL interact with one another. While both have been argued to be powerful and important mechanisms for children and adults, it is necessary to examine their usefulness in more natural and complex scenarios to evaluate this claim. Moreover, their interaction exemplifies how words can be learned based on mechanisms that are fundamentally different in nature: While CSWL is clearly a bottom-up associative process, SLCL exploits top-down knowledge. When conducting CSWL, concurrent linguistic and visual stimuli are mentally associated with each other and the strength of these links is then increased or decreased depending on subsequent evidence or counter-evidence across episodes. That means that no prior knowledge about linguistic or non-linguistic regularities is needed per se; all that is necessary is the multi-modal (linguistic and visual) context and a cognitive system that is able to record co-occurrences between the two. To learn via SLCL, in contrast, language novices need to integrate their knowledge about both linguistic structures and the world. Moreover, SLCL potentially provides more informative and reliable information than CSWL: As demonstrated repeatedly in sentence-processing studies employing the Visual World Paradigm, a visual referent can be immediately identified based on a verb's semantic-syntactic restrictions (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Altmann & Kamide, 1999). Presumably, this information can also be used to identify a referent as the correct mapping for a novel noun. The novel noun *dax*, for instance may instantly be understood as referring to a carrot given the sentence *The man eats the dax.* together with a scene depicting a carrot as the only edible object. That means that SLCL can be *perfectly disambiguating*. On the contrary, the cue that CSWL is based on, that is co-occurrence frequencies of visual referents and spoken words, is inherently indirect and disambiguating only across situations.

In this paper, we examine, whether and under which circumstances, learners apply CSWL and SLCL simultaneously and whether and how both mechanisms modulate each other's effect. Our first motivation driving this investigation is to evaluate the relative importance of both word-learning mechanisms. More importantly though, we consider this interplay as a means of examining the adult word learner's general ability and strategy of deploying multiple mechanisms. In particular, the experiments presented in this paper follow two objectives: firstly, to investigate whether learners are generally able to manage the simultaneous applicability of two mechanisms or whether this complex scenario results in unsuccessful learning. Based on results from previous studies (Gillette et al., 1999; Nappa et al., 2009; Monaghan & Mattock, 2012), we hypothesize learning to be successful (*Multiple-Mechanisms Hypothesis*). Secondly, in case learning is indeed successful, we aim at investigating *how* the two mechanisms are deployed in different learning situations. We hypothesize that the strategy underlying a learner's choice is to first apply the most reliable mechanism and then to decide whether or not to consider a second mechanism (*Prioritize-and-Complement Hypothesis*). In particular, we hypothesize that this decision is made in a situation-dependent way: When the first best mechanism provides sufficient information for disambiguation (that is when the second one would be redundant), the learner will stop the search for cues and mechanisms. In case, it does not, she will look for further available ways to complement her knowledge (Gillette et al., 1999). When this complementary information is not in line with the solution suggested by the first applied mechanism, however, there will be a conflict that results in competition. The knowledge about the reliability of cues is experience-based and generally exploited by the learner (such as word order being a prioritized cue for English native speakers, Nappa et al., 2009).

Our hypotheses regarding the interplay of SLCL and CSWL are based on the assumption that, in any given situation, SLCL will be generally deemed more reliable than CSWL, as outlined above. We thus hypothesize that only SLCL but not CSWL will be employed in case SLCL is perfectly disambiguating (and CSWL would therefore be redundant). Only if SLCL does not provide sufficient information for disambiguation, will CSWL be used. In case both do provide complementary and matching information, they will jointly guide the learner to the meaning of the novel word. In case both are in conflict, SLCL and CSWL will compete.

To investigate these hypotheses in the three scenarios - namely, SLCL and CSWL as complementary, redundant, and conflicting - we exploit a learning procedure for teaching German adults a miniature semi-natural language in a stepwise procedure. In particular, there were three main phases in our experiments: Participants were first familiarized with a set of verbs, then they learned novel nouns, and, finally, noun knowledge was explicitly assessed. Crucially, we used a novel design for noun learning with nouns embedded in linguistic contexts and linguistic contexts situated in visual environments, resembling typical visual-world study trials. In particular, for each noun-learning trial, one spoken subject-verb-object (SVO) sentence was paired with a visual scene containing potential referents. Critical nouns were syntactic objects in these sentences. This constitutes a naturalization of the setting used in learning experiments by embedding the potential referents in visual contexts and the unknown nouns in linguistic contexts.

2 Experiment 1

To investigate whether learners are able to make simultaneous use of CSWL and SLCL as complementary mechanisms, Experiment 1 considers the case where verbal restrictions potentially reduce the set of referents for CSWL. In particular, we manipulated the degree of *verb restriction* (i.e., the degree of disambiguation) within three levels: In condition *Restrictive-1* (*R1*), the restrictive verb (SLCL) identified exactly *one* referent in the scene, which means that using SLCL was both possible and perfectly disambiguating. CSWL, as well, was possible and perfectly disambiguating across situations. In condition *Restrictive-2* (*R2*), the restrictive verb (SLCL) identified *two* potential referents in the scene: While applying SLCL was possible, it was not perfectly disambiguating and CSWL was needed in addition to learn this group of nouns. Alternatively, learning was also possible based on CSWL alone. In condition *Non-restrictive* (*N*), the non-restrictive verb did not constrain the semantic category of the referent, thus leaving four potential targets. That means that SLCL was not possible at all and learning could only succeed based on CSWL.

The Multiple-Mechanisms Hypothesis predicts learning to be successful not only in Condition N but also in Conditions R1 and R2, where both SLCL and CSWL are available. The Prioritize-and-Complement Hypothesis predicts that learning in R1 will entirely rely on SLCL whereas learning in R2 will operate based on both SLCL and CSWL.

2.1 Methods

2.1.1 Participants

50 native speakers of German, mainly students, participated in Experiment 1 for a compensation of €5. 18 learners had to be excluded due to poor verb learning or technical problems, such that only data from 32 participants was included in the analysis (23 female, 9 male, aged between 19 and 31, average age 24).

2.1.2 Materials & Procedure

The aim of the experiment was to teach participants a semi-natural miniature language (modified Indonesian, 'Artonesian'). The language comprised six verbs, 14 nouns, and the article *si* ('the'), all words being (modified or non-modified) Indonesian words. Two of the verbs were non-restrictive (e.g., *tambamema*, 'take') and four were restrictive. Depending on list, participants learned either two restrictive food verbs (e.g., *bermamema*, 'eat') and two restrictive clothing verbs (e.g., *felimema*, 'iron') or two restrictive food verbs and two restrictive container verbs (e.g., *mautimema*, 'empty'). There were two subject nouns (*laki*, 'man' and *gadis*, 'woman') and twelve object nouns (four food items, four clothing items, and four container items). As in Indonesian, word order was SVO.

The experiment proceeded in five stages: verb learning and testing, eye-tracker calibration, and verb repetition (Phase 1); sentence comprehension (learning of six nouns) Block 1 (Phase 2); vocabulary test Block 1 (and verb repetition) (Phase 3); sentence comprehension (learning of six other nouns) Block 2 (Phase 4); vocabulary test Block 2 (Phase 5). Accuracy in the vocabulary tests served as the offline measure of learning success (learning rates and confidence ratings). Additionally, eye-movements were recorded and used as a dependent measure (proportions of looks to objects and characters in the scenes) during noun learning (Phases 2 and 4). The experiment took approximately 30 minutes to complete.

Verb learning In Phase 1, participants were familiarized with a set of verbs: They were presented with simple animated depictions of actions, together with the corresponding spoken verb, and were explicitly asked to memorize the actions' names. Each action-verb pair was named ten times. Then, knowledge of verbs was assessed by presenting a picture and asking learners to pronounce the matching verb. Feedback was provided. The eye-tracker was calibrated and verbs together with depictions were presented and had to be named again. The process of verb learning was not the subject of investigation in itself, but rather served as a pre-requisite for noun learning based on verbal constraints. Only participants who could name the verbs for all six actions were included into noun-learning analyses.

——————————————Insert Figure 2 about here ——————————————-

Noun learning In Phases 2 and 4, scene-sentence pairs were presented (see Fig. 2). Images always depicted one character and four objects embedded in a simple indoor scene. One of the objects was the *target object* (i.e., the referent for the object noun). The others were *competitors* (potential referents matching the verb's constraints) and *distractors* (objects which were neither targets nor competitors). While the background also included objects which could theoretically been considered as referents, they were less prominent in terms of color and visibility. Moreover, the foreground objects were made particularly prominent since they were repeated eight times unchanged across trials.

Items were manipulated according to one three-level within-participant factor (Verb Restriction) such that there were four object nouns per condition (two per condition per block). Four object nouns to be learned (two per block) belonged to the *Restrictive-1 Condition* (*R1*): Verbs preceding these object nouns were restrictive (e.g., *eat*) and there was only one referent in the scene which belonged to the corresponding semantic category (e.g., food), which means that there was no competitor. Another set of four object nouns belonged to the *Restrictive-2 Condition* (*R2*): Verbs were also restrictive but two matching objects were depicted (i.e., there was one competitor). The four object nouns belonging to the *Non-Restrictive Condition* (*N*) occurred in sentences with non-restrictive verbs which means that all four objects were potential referents (i.e., three competitors). That means that learning via SLCL alone was only possible in Condition R1. In Condition R2, SLCL could be applied but could only result in learning when combined with CSWL. In Condition N, SLCL was not available at all. CSWL was possible and perfectly disambiguating across trials in all three conditions.

In Block 1 (Phase 2), no target ever appeared as a competitor or distractor for another noun to ensure participants could not exclude potential referents based on other learned words. In Block 2 (Phase 4), however, learning was potentially simplified by the possibility to exclude those objects as potential referents which were already associated with another noun (principle of *mutual exclusivity*, Markman & Wachtel, 1988). The order of presentation of the 48 trials (24 per Block) was randomized, with each noun repeated four times, and nouns from the three conditions intermixed.

There were eight lists to enable counterbalancing and avoid potential confounds: Firstly, there were two different world-word mappings to avoid that learning effects could be due to inherent associations between words and referents. Secondly, we used two different assignments for objects to conditions (e.g, in half of the lists food objects were in Condition R1 and for the other half they were in Condition R2). Finally, there were two versions of lists that depended on the assignment of items to Blocks 1 and 2. The arrangement of characters and objects in the picture was also counterbalanced, such that targets and distractors were in each position (left-right, bottom-top) equally often. We additionally controlled the objects' spatial relation to the character in the scene to avoid artifacts due to the proximity of agents and objects (participants could, for instance, be biased to believe the object close to the agent must be involved in a taking event). Participants were told that sentences were of the causative SVO form 'someone VERBs something'. They were asked to understand the sentences, and the instructions mentioned that this required learning the unknown words.

Noun testing In each forced-choice vocabulary test trial (Phases 3 and 5), there were six depictions presented on the screen and participants were asked to click on the picture that matched a spoken noun. The scene contained the target and another instance of the target's semantic category, two objects of one of the two other categories, and two characters. In addition to recording the mouse clicks to the chosen referent, we introduced a confidence rating: Participants were encouraged to press a number between 1 (not confident at all) and 9 (very confident) on the keyboard in order to indicate how sure they were about their choice of a referent. Given that we only included confidence ratings into analyses which belonged to correct responses (i.e., when the forced choice was accurate), this measurement provides a more sensitive assessment, reflecting a more graded notion of noun learning. We considered this dependent variable as potentially important to detect differences between conditions because of possible ceiling effects and small amount of data per condition (there were only four nouns to be learned per condition).

2.1.3 Predictions

The Multiple-Mechanisms Hypothesis asserts that both SLCL (R1, R2) and CSWL (R2, N) will be applied predicting learning rates in the vocabulary test to be above chance in all conditions. The Prioritize-and-Complement Hypothesis predicts learning will rely on SLCL alone in Condition R1 but will operate on both SLCL and CSWL in Condition R2: In R1, SLCL is perfectly disambiguating whereas in Condition R2 it is not. Applying CSWL in Condition R2 predicts learning to be above chance - without CSWL it could not be successful. Applying SLCL in Condition R1 and (in addition to CSLW) in Condition R2 predicts, firstly, learning to differ from learning in Condition N where only CSWL is applicable; secondly, eye movements are predicted to reflect a preference for the referents matching the verbal constraints during the object noun is spoken (i.e., during the second, postverbal nominal phrase, NP2).

2.2 Data analysis & Results

Off-line Results (Phases 3 and 5) Assuming that learners quickly understand that the second noun phrase (NP2) refers to an inanimate object, we compared noun learning to a chance level of 25% since noun-test pictures contained four such objects. Noun learning was reliably better than chance across all conditions (72%, see Table 1) and highly correlated positively with all confidence ratings, including those for wrong choices ($r_s = .430, p < .001$). Learning was significantly better in Block 2 than Block 1 (all conditions: $\chi(1) = 30.769, p < .001$; Condition R1: $\chi(1) = 6.309, p < .05$; Condition R2: $\chi(1) = 10.167, p < .01$; Condition N: $\chi(1) = 16.568, p < .001$; see Table 1). Likewise, confidence ratings were higher in Block 2 than Block 1 (all conditions: $\chi(1) = 12.849, p < .001$; Condition R1: $\chi(1) = 5.416, p < .05$; Condition R2: $\chi(1) = 5.476, p < .05$; Condition N: $\chi(1) = 10.688, p < .01$; see Table 2).

——————————Insert Table 1 about here ——————————-

——————————Insert Table 2 about here ——————————-

Descriptively, nouns were learned best and confidence was rated highest in Condition R1 and worst in Condition N (see Fig. 3). Both learning rates and confidence ratings (of learned nouns), collapsed over blocks, were analyzed using *linear mixed-effects* (*lmer*) models, using logistic regression for the categorical learning rates (logit link function, from the lme4 package in R, Bates, 2005) and linear regression for the continuous confidence ratings, with participants and items as random factors. To determine whether the fixed factor (Verb Restriction) had a main effect (i.e., whether including the factor significantly improved the predictive power of the model) we compared the models that include and exclude this factor with a chi-square test (Baayen, Davidson & Bates, 2008). Contrasts between levels of a factor (R1, R2, N) were assessed by the ratio of regression coefficients (β) and standard errors (SE) since the p-values produced by *lmer* (Wald z test) may be anti-conservative (Baayen et al., 2008): If the coefficient was larger than twice the standard error, the difference was considered to be significant. Tables of these statistical comparisons are provided below. The formulas describing the *lmer* models are of the following form: dependent variable (learning rate/confidence rating) is a function of ($\sim$) the independent variable (Verb Restriction) plus random effects (subjects and items). For confidence ratings we additionally calculated Monte

Carlo Markov Chain values (MCMCs) whose p-values are a good estimate of a factor's significance (but are only applicable for continuous variables, Baayen et al., 2008).

For confidence ratings, we found a main effect of factor Verb Restriction ($\chi(2) = 31.833, p < .001$) as well as significant differences between all conditions: R1 and R2 (Row 2), R1 and N (Row 3), as well as R2 and N (Row 6) (Table 3). While learning rates showed the same general pattern, there was no significant main effect ($\chi(2) = 4.289, p = .117$).

——————————Insert Figure 3 about here ——————————-

——————————Insert Table 3 about here ——————————-

Eye movements For eye-movement analysis, we examined whether there was at least one starting inspection in a trial for each area of interest (AOI; the character and the four objects) during the NP2 time window. An inspection was defined as two or more consecutive fixations to the same AOI. This measurement has been used in the visual world paradigm (Knoeferle et al., 2005; Knoeferle & Crocker, 2007) with the rationale of filtering very short fixations that may have little to do with the task. We counted only those inspections that started after the relevant linguistic information had been heard in order to restrict analyses to looks driven by a linguistically motivated change of attention. The analyzed NP2 time window started 200ms after onset of NP2 (because in that time only the non-differentiating determiner was heard) and ended at the offset of the noun. Data was analyzed as learning rates, using linear mixed-effects models with logistic regression.

——————————Insert Figure 4 about here ——————————-

There was a main effect of AOI ($\chi(4) = 280.910, p < .001$) as well as an interaction of AOI with Verb Restriction ($\chi(10) = 50.574, p < .001$). This interaction was caused by clear differences between conditions for inspection patterns (see Fig. 4): In Condition R1, the target object was inspected reliably more often than the character and the distractor objects (main effect AOI: $\chi(4) = 69.236, p < .001$; Table 4, Rows 2-5; Fig. 4). All objects, however, were inspected more frequently than the character (Table 4, Rows 12-15). In Condition R2, the target was inspected significantly more often than the other objects (main effect AOI: $\chi(4) = 118.750, p < .001$; Table 4, Rows 17-20). Moreover, the competitor was looked at reliably more often than the other objects (except the target) (Table 4, Rows 22-25). No other objects than target and competitor were looked at significantly more often than the character (Table 4, Rows 27-30). For Condition N, the target was inspected reliably more often than the character and the distractors (main effect AOI: $\chi(4) = 138.650, p < .001$; Table 4, Rows 32-35), except that the difference between looks to the target and to one of the three competitors was non-significant (Tables 4, Row 33). We found in a posthoc analysis that this distractor was exactly that one which shared the semantic category with the target (e.g., when the target was bucket, the distractor was another container). There were also significantly more looks to this distractor than to the other distractors and the character (Table 4, Rows 37-40). All objects were looked at significantly more often than the character (Table 4, Rows 42-45).

——————————Insert Table 4 about here ——————————-

To examine whether verbal constraints had an immediate influence on learners' attention, we moreover analyzed eye-movements during the verb region (see Fig. 5). We found a main effect of AOI for all levels of verb restriction (Condition R1: $\chi(4) = 20.374, p < .001$, Condition R2: $\chi(4) = 15.082, p < .010$, Condition N: $\chi(4) = 36.544, p < .001$). In Condition R1, the target object was looked at reliably more often than the rest (Table 5, Rows 2-5). In Condition R2, the target object was inspected significantly more often than the character but not the distractors (Table 5, Rows 17-20). While the difference between target and competitor was not significant, the competitor was inspected significantly more often than the character and the distractors (Table 5, Rows 23-25). Importantly, when comparing inspections to the two objects matching the verbal restrictions (i.e., target plus competitor) versus the two distractors, looks to the matching group were significantly more frequent than looks to the distractor group ($\beta = 0.416, SE = 0.156, z = 2.670, p < .01$).

In Condition N, the target object was looked at reliably more often than the other regions except the distractor of the target's semantic category (di1; Table 5, Rows 32-35). This distractor was looked at reliably more often than the character (Table 5, Row 32).

——————————Insert Figure 5 about here ——————————-

——————————Insert Table 5 about here ——————————-

2.3 Discussion

To summarize, learning rates were above chance for all three levels of verb restriction. Moreover, we found significant differences in confidence ratings between levels of verb restriction: highest ratings in Condition R1 and lowest in Condition N. This pattern was also numerically present in the learning rates and both measurements were highly positively correlated. Eye-movements during NP2 further revealed different patterns for the three conditions: In R1, we found a clear preference for inspections to the target object; differences between looks to the target and looks to all other AOIs were significant. All objects, however, were looked at more often than the character. In R2, in contrast, learners looked most often at the target object while additionally having a secondary significant preference for inspecting the competitor object. No object other than target and competitor was inspected more often than the character. In N, the target object and the distractor object which belonged to the same semantic category as the target were looked at most frequently. While both were looked at significantly more often than the other AOIs, the difference between the two was not significant. As in R1, all objects were inspected more frequently than the character. During verb region, there was a reliable preference to inspect the object matching the verbal constraints in Condition R1. In R2, the differences between looks to target and distractors lacked significance; however, target and competitor together were inspected significantly more often than the two distractors together. In Condition N, during NP2, the target object and the distractor object which belonged to the same semantic category as the target were looked at reliably most frequently.

The finding that learning rates were above chance for all levels of verb restriction clearly supports the Multiple-Mechanisms Hypothesis: Participants learned successfully in the complex scenario of multiple applicable word-learning mechanisms. Moreover, we found clear evidence for the Prioritize-and-Complement Hypothesis since, in Condition R2, learners applied both SLCL and CSWL: CSWL must have been used because otherwise learning rates could not be above chance. At the same time, eye-movements reflect a significant preference for the two objects matching the verbal constraints, suggesting that SLCL has been applied. The use of SLCL in Condition R2 is additionally supported by the finding that confidence ratings were significantly higher than in Condition N, where only CSWL was applied. The results also suggest that in Condition R1, where SLCL was perfectly disambiguating, learning relied only on this mechanism, as predicted by the Prioritize-and-Complement Hypothesis. Eye-movements reflect a clear preference for the sole object matching the verbal constraint (already from verb region on). To be confident that CSWL has not been applied in the case of SLCL being perfectly disambiguating, however, we conducted Experiment 2.

Eye-movements during verb region in Conditions R1 and R2 reveal that verbal constraints had an immediate effect on learners' attention: The objects matching the verbal constraints were preferred over the other objects.

The gaze pattern for Condition N in both verb region and NP2 (preferences for the target as well as for the distractor which shared its category) is unforeseen but potentially interesting: Although the non-restrictive verbs (e.g., *tambamema*, 'to take') did not semantically constrain the category of the object referent, the distractor with the same category as the target (e.g. container objects) was preferred over the other distractors (which were of categories associated with other restrictive verbs). A possible explanation is that learners simply became very sensitive to categories during the experiment, attending more to distractors of the same category as the target than to other distractors. Alternatively, participants may have learned a new co-occurrence restriction for non-restrictive verbs during noun learning (e.g., container objects and 'take') simply by recognizing the co-occurrence of verbs and categories and, probably, by excluding the objects which were already selected by another verb.

Eye-movements during the verb region and NP2 generally reveal that the objects that matched the verbal constraints according to the way predicted were attended to more strongly than the other objects. That means that our assumptions about the restrictive categories were supported (iron, sew - t-shirt, trousers, skirt, jacket; eat, barbecue - broccoli, chicken, tomato, toast; fill, empty - vase, watering can, bottle, bucket).

3 Experiment 2

Experiment 1 demonstrated that SLCL and CSWL can be used simultaneously to complement one another, supporting both the Multiple-Mechanisms Hypothesis and the Prioritize-and-Complement Hypothesis. It remains unclear, however, if learners would still apply both mechanisms when SLCL is sufficient independently (i.e., the use of CSWL is redundant). Applying both in this situation would contradict the Prioritize-and-Complement Hypothesis. While eye-movements in Condition R1 in Experiment 1 suggest that CSWL was not used when SLCL was perfectly disambiguating, the data remain inconclusive with respect to this issue. To address this question, noun learning in Experiment 2 was designed according to two manipulations. Firstly, each noun had two potential meanings, one with a higher co-occurrence frequency (83%), and one with a lower co-occurrence frequency (50%), but both with a co-occurrence frequency higher than the one of the distractors (17%). Secondly, each noun occurred in sentences with one of two levels of verb restriction: In Condition *R(estrictive)*, nouns followed restrictive verbs, in Condition *N(on-restrictive)*, they followed non-restrictive verbs. That means that while in Condition R, both cross-situational statistics and sentence-level constraints were available, in Condition N, only CSWL could be applied. Crucially, in Condition R, both CSWL and SLCL supported the 83% referent, which means that using either mechanism was sufficient.

3.1 Methods

3.1.1 Participants

29 German native speakers, mainly students (from different disciplines), took part in Experiment 2 for a reimbursement of €7.50. Five of them had to be excluded due to verb learning problems. Data of the remaining 24 learners (5 males, 19 females, mean age 24) entered analyses. None of these participants had participated in Experiment 1.

3.1.2 Materials & Procedure

The experimental materials and procedure were similar to those in Experiment 1. The miniature language consisted of 18 nouns (the same two character names as in Experiment 1 and 16 object names), four verbs (two restrictive, e.g. *bermamema*, 'eat', two non-restrictive, e.g. *tambamema*, 'take'), and the article *si* ('the').

The experiment consisted of the following phases: verb learning, verb testing, eye-tracker preparation and verb repetition (Phase 1), noun learning Block 1 (eight new object nouns, Phase 2), vocabulary test Block 1 (Phase 3), noun learning Block 2 (another eight object nouns, Phase 4), vocabulary test Block 2 (Phase 5), and a final verb check. Eye-movements were recorded in Phases 2 and 4. The experiment lasted approximately 45 minutes.

Verb learning In Phase 1, participants were familiarized with the verbs. Verb Training 1 consisted of a subset of the materials of Experiment 1 and the procedure was similar. Participants were presented spoken verbs and simple animations concurrently. Instructions explicitly told them that the verbs are the names of the depicted actions and that their task is to learn these mappings. Each verb was repeated nine times (with two different animations per verb). To further facilitate verb familiarization, there was an additional forced-choice training (Verb Training 2). Pictures of the four actions were visible at the same time (the last position of the animations), one spoken verb was presented, and participants were requested to click onto the action matching the verb (each verb was tested twice). The verb testing was identical to that of Experiment 1. Only participants who chose the correct actions for all four verbs in this test were included into analyses. The eye-tracker was calibrated and the verbs were quickly repeated. Before learners were introduced into Phase 2, there was a short familiarization with the pictures of all objects before the noun-learning phase started, in order to assure recognizability.

——————————Insert Table 6 about here ——————————-

Noun learning During noun learning (Phases 2 and 4), participants were exposed to pairs of static scenes and spoken SVO sentences (see Table 6), as in Experiment 1. Participants were asked to understand the sentences but were not explicitly told to learn the novel nouns. Noun learning consisted of 96 scene-sentence pairs, six presentations for each of the 16 object nouns. Eight object names were trained in Block 1 and eight in Block 2, such that each block consisted of 48 trials.

Each noun had two potential referents (i.e., meanings), one co-occurred with the noun in 83% of the trials (the high-frequency object, e.g., corn and hamburger in Table 6) and one co-occurred with the noun in 50% of the trials (the low-frequency object, e.g., socks and jumper in Table 6). Objects other than the high-frequency object and the low-frequency object (i.e., distractors) all had a co-occurrence frequency of 17% with a noun. To avoid learning strategies based on mutual exclusivity, no object belonged to the 83% group for one noun and to the 50% group for another noun (that is why there were twice as many objects as nouns). In addition to the different frequency groups, nouns were in one of two conditions (factor Verb Restriction): The eight object nouns that were in Condition *N(on-restrictive)* always occurred with a non-restrictive verb. The eight object nouns in Condition *R(estrictive)* occurred with a restrictive verb in 83% of their presentations (five of six). Importantly, in these restrictive trials, there was only one object depicted that matched the verbal restrictions. In Condition R, both verbs (=SLCL) and co-occurrence frequencies (=CSWL) supported the 83% meaning (e.g., hamburger, Table 6). Given that nouns' 83% referents were sometimes not depicted (i.e., in 1/6 trials), referential uncertainty was made more complex for the learner and at the same time more naturalistic: As in the real world, not everything in the scene was always mentioned and not everything mentioned was always in the scene.

The manipulations described above necessitated three trial-type categories: In trials of Category

1 (three of six trials), both the high-frequency object and the low-frequency object were contained in the scene. In trials of Category 2 (two of six trials), the low-frequency object was not included in the scene, while scenes belonging to trials of Category 3 (one of six trials) depicted neither the high-frequency nor the low-frequency referent. All scenes contained four objects. In trials of Category 1 and 2, there was always only one object matching the verbal constraints (in Condition R). In four of six trials, the character was not included in the scene in order to implicitly remind participants of the fact that not everything that is mentioned in the sentence needs to be depicted.

Both Block 1 (Phase 2) and Block 2 (Phase 4) were subdivided into two parts for presentation. Each subpart consisted of 24 trials (four novel nouns). The order of presentation within these subparts was randomized. There was no vocabulary test or brake between the two subparts. We used this control in order to facilitate learning of the (compared to Experiment 1) enlarged set of 18 new nouns. Pictures were counterbalanced in the same way as in Experiment 1. Each of the 36 objects was presented 12 times, independent of condition or group, to avoid visual dominance effects.

Noun testing For the vocabulary tests in Phases 3 and 5, learners heard a noun and were asked to indicate which one of four visual objects was the referent, by clicking on it. There were two different test types (see Table 7 for example trials). Either the 83% object, the 50% object, and two distractors were depicted (Test Type 1) or the 50% object and three distractors were depicted (Test Type 2). There were 32 test trials (16 per test type), each object name was used twice, once in each trial type, respectively.

Participants were presented training and testing trials according to one of four lists: Firstly, nouns were assigned to one of the two conditions. Secondly, objects were assigned to either the 83% group or the 50% group.

——————————Insert Table 7 about here ——————————-

3.1.3 Predictions

The Multiple-Mechanisms Hypothesis predicts learning to be above chance in both Conditions N and R. The Prioritize-and-Complement Hypothesis predicts for Condition N that CSWL will be applied because it is the only available mechanism. The further hypothesis that CSWL operates in a parallel and probabilistic manner (Vouloumanos, 2009), predicts that learners should be sensitive to the differences between the three co-occurrence frequencies 17% (distractors), 50% (low-frequency candidate), and 83% (high-frequency candidate). In particular, in Test Type 1, learners should prefer to choose the 83% object rather than the 50% object and the 17% distractors. While the 50% referent may moreover be chosen more often than the 17% distractors, the presence of the 83% object may suppress this difference. In Test Type 2, however, the 50% referent is clearly predicted to be chosen more often than the 17% distractors. For Condition R, the Prioritize-and-Complement Hypothesis however predicts that learning will operate only on SLCL since it is perfectly disambiguating and using CSWL would therefore be redundant. This predicts learners to have no sensitivity for the difference between the 50% object and the 17% objects. Therefore, they should show no preference for the 50% object over the 17% objects in neither test type.

Concerning eye movements in Condition R, the Prioritize-and-Complement Hypothesis predicts a clear preference for inspecting the 83% object, beginning in the verb region and, more pronounced, during NP2; there should be no secondary preference for the 50% object during NP2 as would be predicted by a theory assuming that CSWL does operate in addition to SLCL. For Condition N, this secondary preference for inspecting the 50% object is expected.

3.2 Data analysis & Results

Off-line Results Learning rates (83%-candidate chosen in Test Type 1) were significantly above chance (25%) for both conditions (N: 60%, $t(23) = 7.995, p < .001$; R: 84%, $t(23) = 16.284, p < .001$). Learning was significantly better in Condition R than Condition N ($\chi(1) = 15.122, p < .001$).

For analyzing differences in the frequencies of decisions for the four depicted objects in the vocabulary test (and how these differences differed in conditions), we conducted repeated measures ANOVAs. The chosen meaning served as the dependent variable (Test Type 1: 83%, 50%, Distractor 1, Distractor 2; Test Type 2: 50%, Distractor 1, Distractor 2, Distractor 3). Data was averaged across items (F_1) and subjects (F_2), for both levels of verb restriction (N and R) and both test types (Test Type 1 and 2). [1]

First, data from test trials of Test Type 1 (83% candidate and 50% candidate present) was analyzed (see Fig. 6). Main effects were found for both conditions (N: $F_1(3, 69) = 40.272, p < .001; F_2(3, 45) = 51.277, p < .001$; R: $F1(3, 69) = 225.448, p < .001; F_2(3, 45) = 304.856, p < .001$): The 83% object was chosen significantly more often than each other object in both conditions (see Table 10; N: Rows 1-3; R: Rows 13-15).

Moreover, when entering the data of only 83% choices and 50% choices into linear mixed effect models (using logistic regression), we found Verb Restriction (N/R) to be a predictor for Chosen Meaning (83% referent/50% referent; $\chi(1) = 17.300, p < .001$): The high-frequency object was chosen reliably more often in Condition R than in Condition N ($\chi(1) = 28.420, p < .001$, Table 8, Row 4) and the low-frequency object was picked reliably more often in Condition N than in Condition R ($\chi(1) = 12.688, p < .001$, Table 8, Row 6).

————————————Insert Table 8 about here ————————————-

————————————Insert Table 9 about here ————————————-

————————————Insert Figure 6 about here ————————————-

For Test Type 2 (only 50%-object available; Fig. 6), we again conducted one-way ANOVAS with Chosen Meaning as fixed factor, separately for conditions and found that Chosen Meaning had a significant effect for both levels of verb restriction (N: $F_1(3, 69) = 9.938, p < .001; F_2(3, 45) = 9.018, p < .001$; R: $F1(3, 69) = 15.165, p < .001; F_2(3, 45) = 22.132, p < .001$). Pairwise comparisons reveal notable differences between conditions. In Condition R, one distractor was chosen

[1] The reason for not using linear mixed effect models was that multilevel logistic regression, which would have been required due to the fact that Chosen Meaning has four levels, was not implemented for the analysis software we were using (statistical package R).

significantly more often than each other object (Table 11: Rows 13-15); this distractor shared the semantic category with the non-present 83% referent (from now on referred to as *category associate* or *CA*). No other difference was significant (Table 11, Rows 16-24). In Condition N, in contrast, both the category associate and the 50% object were chosen significantly more often than the two distractor objects (Table 11 Rows 2-3, 5-6). The difference between category associate choices and 50%-object choices was not significant (Table 11, Row 1).

When entering the data of only CA choices and 50% choices into linear mixed effect models using logistic regression, we found that condition predicts whether the low-frequency object or the category associate is chosen ($\chi(1) = 15.651, p < .001$, Table 9, Row 2): Learners made significantly more category-associate decisions in Condition R than in Condition N ($\chi(1) = 7.008, p < .01$, Table 9, Row 4) and significantly more 50%-object choices in Condition N than in Condition R ($\chi(1) = 17.612, p < .001$, Table 9, Row 6).

——————————Insert Table 10 about here ——————————-

——————————Insert Table 11 about here ——————————-

Eye movements Eye movements during noun learning were analyzed for regions verb and NP2 of trials belonging to Trial Category 1 (which contained scenes depicting both the high-frequency object and the low-frequency object), separately for both conditions. Data was treated as in Experiment 1.

——————————Insert Figure 7 about here ——————————-

For Trial Category 1 (high-frequency object and low-frequency object included in the scene), we found effects of Area of Interest (AOI) in verb region of Condition R ($\chi(3) = 108.700, p < .001$, Table 12, Rows 10-12) but not in verb region of Condition N ($\chi(3) = 6.258, p = .100$, Table 12, Rows 2-4) and in NP2 of both conditions (N: $\chi(3) = 13.463, p < .01$, Table 12 , Rows 6-8; R:

$\chi(3) = 97.157, p < .001$, Table 12, Rows 14-16): The high-frequency object was inspected reliably more often than any other object (i.e., the low-frequency object and both distractors). There was an interaction between AOI and Verb Restriction for both verb region ($\chi(4) = 40.48, p < .001$) and NP2 ($\chi(3) = 28.366, p < .001$; Fig. 8). Crucially, this interaction is still present when only looks to the high-frequency object and looks to the low-frequency object are included (verb: $\chi(1) = 30.529, p < .001$; NP2: $\chi(1) = 11.026, p < .001$): The difference between looks to both objects was reliably larger in Condition R than Condition N. A further analysis reveals that learners made significantly more looks to the low-frequency object during NP2 in Condition N than in Condition R ($\chi(1) = 62.615, p < .001$, Table 13). See Fig. 7 for time graphs.

——————————Insert Figure 8 about here ——————————-

——————————Insert Table 12 about here ——————————-

——————————Insert Table 13 about here ——————————-

3.3 Discussion

First of all, results from Experiment 2 provide further support for the Multiple-Mechanisms Hypothesis: Learning was above chance even when more than one word learning mechanism was applicable (Condition R). Moreover, the data support the hypothesis that CSWL operates in a parallel and probabilistic manner: In Condition N, when CSWL was the only mechanism that was applicable, learners were sensitive to the differences between co-occurrence frequencies of 83%, 50%, and 17%. In Test Type 1, the 83% candidate was preferred over the other objects. While the 50% object was not chosen reliably more often than the 17% distractors, such a trend was found, supported by the finding that the 50% object was selected significantly more often in Condition N than Condition R. In Test Type 2 the sensitivity for the difference between co-occurrence rates of 17% and 50% was clearly reflected in learners' decisions. This result is in line with Vouloumanos

(2008) and all of our three main hypotheses.

Interestingly, this sensitivity was completely blocked in Condition R, when SLCL was applicable: Learners did not choose the 50% object more often than the 17% distractors in either test type. This result clearly supports the Prioritize-and-Complement Hypothesis: CSWL was not employed because SLCL was perfectly disambiguating.

With regard to the high probability of choosing the referent that shared its semantic category with the non-present 83% referent (the category associate) for both levels of verb restriction, we suggest this was likely the result of a general (possibly experiment induced) strategy of selecting an object which is semantically closest to the non-present 83% referent. The finding that the 83% object was chosen significantly more often in Condition R than Condition N supports, again, that verbal restrictions (SLCL) result in better learning than CSWL. This finding is in line with results from Experiment 1 (Conditions R1 versus N).

It is possible that learners' choices in the vocabulary test were influenced by the way the experimental instructions were worded. In particular, the fact that participants were asked to *understand* the sentences rather than to *learn* the nouns might have biased them to believe that an object similar in type as the assumed referent is a better choice than considering co-occurrence frequencies. However, this strategy would not explain the difference between selection preferences in Conditions N and R: Besides the preference for the category associate in both conditions, there is a preference to choose the 50% object only in Condition N.

Consistent with the off-line results, eye-movements support the sensitivity for fine-grained differences in co-occurrences to be present in Condition N but blocked in Condition R, supporting the Prioritize-and-Complement Hypothesis: While there was a very clear preference during learning trials of Category 1 (83% object and 50% object depicted) to inspect the high-frequency candidate in Condition R, during the verb and NP2, both the high-frequency object and low-frequency object were considered during NP2 in Condition N (see time graphs in Fig. 7). This pattern reveals that CSWL was not applied in Condition R.

4 Experiment 3

Both Experiments 1 and 2 are in line with the Multiple-Mechanisms Hypothesis and the Prioritize-and-Complement Hypothesis. It is possible, however, that neither is valid in a situation in which the two applicable mechanisms are in conflict. In Experiment 3, we evaluate the interaction of CSWL and SLCL in this scenario.

Each noun in Experiment 3 had two potential meanings, as in Experiment 2: One referent co-occurred with the noun in 83% of the noun's usages, the other referent co-occurred with the noun in only 50% of the cases. Additionally, each noun belonged to one of two conditions: In Condition ***N****(on restrictive)*, a noun was always preceded by a non-restrictive verb such as *take*. In Condition ***R****(estrictive)*, a noun was preceded by a restrictive verb such as *eat* half of the time. While CSWL supported the 83% object in both conditions, SLCL (only applicable in Condition R) supported the 50% referent. Importantly, those trials in Condition R, that contained a restrictive verb, did not contain an object matching these verbal constraints. Only in the following trial, the matching object was presented (together with a non-restrictive verb). That means that in no single trial but only across trials, SLCL provided perfect disambiguation. The Prioritize-and-Complement Hypothesis therefore predicts that learning in Condition R will operate on both SLCL and, to complement, CSWL. Since both provide contradicting cues, a situation of competition is expected.

4.1 Methods

4.1.1 Participants

28 native speakers of German who had not participated in either of Experiments 1 or 2 took part in Experiment 3, four of which had to be excluded due to unsuccessful verb learning. Data of 24 participants was analyzed (4 males, 20 females, mean age 24). They received €6 for taking part in the experiment.

4.1.2 Materials & Procedure

As in the first two experiments, Experiment 3 sought to teach participants a miniature semi-natural language. The language was similar to the one in the other experiments. It comprised the same four verbs as Experiment 2, twelve nouns ('man', 'woman', and ten object names), and the article *si*.

The experiment consisted of the following phases: Verb Training 1, Verb Training 2, and verb testing, eye-tracker preparation, and verb repetition (Phase 1), noun learning Block 1 (2 object nouns, Phase 2), vocabulary test Block 1 (Phase 3) and verb repetition, noun learning Block 2 (four other object nouns, Phase 4), vocabulary test Block 2 (Phase 5), noun learning Block 3 (four other object nouns, Phase 6), vocabulary test Block 3 (Phase 7) and final verb test. The reason for splitting noun learning and noun testing into 3 blocks (two nouns, four nouns, four nouns) was to facilitate learning. The entire experimental procedure lasted approximately 40 minutes.

Phase 1 was exactly as in Experiment 2: Participants were introduced to the experiment, verbs were trained and tested, and the eye-tracker was adjusted before pictures of all object were shortly presented.

Noun learning Before noun-learning trials were presented, the nouns for 'man' (*laki*) and 'woman' (*gadis*) were explicitly introduced, together with the depictions used in the experiment. We informed participants that while scenes are often helpful for understanding the sentence, they do not necessarily fully correspond to it. We considered this level of explicitness necessary to avoid confusion given that the two cues supported different visual referents for each noun. Materials for the noun-learning phases (Phases 2, 4, and 6) were basically as in the other experiments: Participants were exposed to pairs of static scenes and spoken SVO-sentences. These sentences consisted of the verbs that had been learned before and novel nouns as subjects and objects. Scenes depicted semi-natural indoor-scenes with characters and objects. Participants were told that their task is to learn the noun meanings.

As in Experiment 2, each of the ten object nouns to be learned had a high-frequency (83%) meaning and a low-frequency (50%) meaning, which means that there were 20 objects. Object nouns occurred in sentences with one of two levels of verb restriction (five with R(estrictive) verbs and five with N(on-restrictive) verbs). The crucial difference to Experiment 2 was that in Experiment 3, in Condition R, the verbal restrictions supported the low-frequency meaning rather than the high-frequency meaning. Therefore, the high-frequency meaning was supported by CSWL and the low-frequency meaning by SLCL. Furthermore, restrictive verbs were used only in that half of the trials in which the scene *did not* include the SLCL-supported referent. That means that restrictive verbs and referents matching the verb's semantic category were never co-present and verb information had to be memorized across trials. We enforced this constraint to avoid trials with direct conflict between SLCL and CSWL, which would potentially confuse learners. We hypothesized that learners would nonetheless be able to use verb information across trials, based on studies by Arunachalam & Waxman (2010a) and Yuan &Fisher (2009) demonstrating this ability in cases of syntactic bootstrapping.

Each noun was presented six times such that there were 60 trials, 12 in Block 1 and 24 each in Blocks 2 and 3. In three of six trials per noun (Trial Category 1), the scenes contained the high-frequency object and the low-frequency object (but never a restrictive verb in neither condition), in two of six trials (Trial Category 2), only the high-frequency object was depicted (and the verb was restrictive in Condition R), and in one of six trials (Trial Category 3), neither the low-frequency referent nor the high-frequency referent were included in the scene (and the verb was restrictive in Condition R, see Table 14).

Participants were presented trials according to one of four lists: There were two world-word-mappings and two assignments of items to conditions. By assigning the items to both conditions across lists, the assignment to object type (low-frequency or high-frequency) was manipulated at the same time. The noun *firel*, for instance, was in Condition R and had a food object as low-frequency referent (and a clothing item as high-frequency object) in two of the lists. In the two remaining lists, it belonged to Condition N and had a clothing item as low-frequency referent (and a food item as high-frequency candidate).

In noun-learning Block 1 (Phase 2), two nouns were introduced (one per condition). Block 2 (Phase 4) and Block 3 (Phase 6) each contained four novel nouns (two per condition). The presentation of trials was pseudo-randomized within blocks with the following constraints: Each trial of Trial Categories 2 and 3 was directly followed by a trial of Trial Category 1. That means that each trial containing a restrictive verb (e.g., eat) was directly followed by a trial containing a referent that matched the verbal constraints (e.g., corn). The reason for this control was to reduce the difficulty of using verb information across trials. The resulting trial pairs were then randomized in order.

——————————Insert Table 14 about here ——————————-

Noun testing As in the other experiments, participants' task in the vocabulary tests (Phases 3, 5, and 7) was to make a forced choice about the meanings of nouns: They were asked to click onto the object depicted on the scene which matched the spoken noun. As opposed to Experiments 1 and 2, all 20 objects were shown for each test trial. [2]

In the vocabulary test in Block 1 (Phase 3), we tested knowledge about the two nouns which participants had been familiarized with in noun learning Block 1 (Phase 2); in the vocabulary test Block 2 (Phase 5), the four nouns presented in noun-learning Block 2 (Phase 4) were assessed; in the vocabulary test Block 3 (Phase 7), participants were tested on their knowledge about the four nouns presented in noun-learning Block 3 (Phase 6), accordingly.

The vocabulary test in Block 3 (Phase 7) was followed by another verb test: Participants were asked to decide in a forced-choice way which verb matched which action depiction. This was done in order to assess whether participants were entirely familiar with all four verbs at the end of the experiment.

4.1.3 Predictions

The Multiple-Mechanisms Hypothesis predicts learning rates to be above chance for both conditions. The Prioritize-and-Complement Hypothesis predicts noun learning in Condition N to be based on the only available mechanism CSWL, predicting a clear preference for the 83% target to be chosen

[2] This modification addresses the criticism by Smith, Smith, & Blythe, 2010, who argue that a constrained set of objects in the test simplifies the task and cannot be used to evaluate pure CSWL.

in the vocabulary test. For Condition R, the Prioritize-and-Complement Hypothesis predicts full competition between CSWL and SLCL: Since SLCL does not provide perfect disambiguation in any trial, learning is predicted to operate on CSWL in addition to SLCL, in order to complement it. This competition should result in a similar number of 83% choices and 50% choices.

4.2 Data analysis & Results

Performance in noun learning (i.e., learning either the low-frequency or the high-frequency meaning) was clearly better than chance (5%): 84.2% for both conditions taken together ($t(23) = 26.319, p < .001$), 87.5% for Condition N ($t(23) = 24.665, p < .001$), and 80.8% in Condition R ($t(23) = 20.206, p < .001$). The effect of factor Verb Restriction on the factor Chosen Meaning in the vocabulary test was analyzed using linear mixed-effects models with logistic regression (logit link function), with participant and item as random factors. Importantly, we found a main effect of Verb Restriction on Chosen Meaning ($\chi(1) = 59.300, p < .001$, Table 15): In N, learners chose the high-frequency meaning 97% of the times and the low frequency meaning only 3%. In Condition R, however, high- and low-frequency meanings were chosen about equally often (high: 48%, low: 52%) (see Fig. 9). The effect of Verb Restriction on Chosen Meaning also confirms that learners were able to use verb information across trials. Importantly, this pattern can be observed across participants, excluding the possibility that one group of learners always chose the 83% referent and others always the 50% referent.

——————————Insert Figure 9 about here ——————————-

——————————Insert Table 15 about here ——————————-

Results from eye movements during learning are not very informative, given that restrictive verbs and the verb-supported 50% object were never co-present. However, the inspection pattern for those trials which depicted both the 50% object and the 83% object (but which contained always non-restrictive verbs), is consistent with the off-line results: There was a clearer preference to inspect the 83% object in Condition N than Condition R, as reflected by a significant interaction between factors Area of Interest (AOI, 50% object versus 83% object) and Verb Restriction (R versus N; $\chi(2) = 16.773, p < .001$).

4.3 Discussion

First of all, Experiment 3 is consistent with both Experiments 1 and 2 in its compatibility with the Multiple-Mechanisms Hypothesis: Learning was successful in both conditions, even when two mechanisms were applicable (Condition R). Importantly, Experiment 3 moreover provides additional support for the Prioritize-and-Complement Hypothesis: As predicted, SLCL and CSWL competed, resulting in an equal number of CSWL-supported and SLCL-supported choices in the vocabulary test. We interpret this competition to be due to the fact that SLCL, as the most reliable applicable mechanism, did not provide perfect disambiguation in any given trial. Therefore, CSWL was additionally used and the learner received contradictory cues about each noun's meaning.

5 General discussion

In three experiments, we evaluated two main hypotheses regarding the possible interplay of word learning mechanisms, the Multiple-Mechanisms Hypothesis and the Prioritize-and-Complement Hypothesis. Concretely, we investigated the interaction of the two mechanisms word learning based on sentence-level constraints (SLCL) and cross-situational word learning (CSWL).

In Experiment 1, we evaluated whether SLCL and CSWL are both used when they complement one another, that is, when SLCL is not perfectly disambiguating and CSWL provides additional cues. We examined learning in three conditions: When only CSWL was possible (Condition N), when either of both SLCL and CSWL was possible and sufficient (i.e., perfectly disambiguating; R1), and when CSWL was sufficient and SLCL was applicable but not perfectly disambiguating and CSWL was needed in addition (R2). We found that learning in all conditions was successful, supporting the Multiple-Mechanisms Hypothesis. The results are moreover in line with the Prioritize-and-Complement Hypothesis: In Condition R2, successful learning, difference of learning rates compared to Condition N, and eye movements reveal that both SLCL and CSWL were simultaneously applied. Learning rates and eye-movements in Condition R1, on the contrary, suggest that learning operated on SLCL only, also confirming our assumption that learners consider SLCL more reliable than CSWL.

In Experiment 2, we investigated how SLCL and CSWL are employed when SLCL is perfectly disambiguating by itself and the mechanisms are therefore redundantly applicable. In particular, we examined whether learning in CSWL works parallel and probabilistic both when SLCL is applicable and when it is not. We found that while learners showed sensitivity to smaller differences in co-occurrence frequencies during CSWL learning alone (Condition N), supporting parallel learning, this sensitivity was blocked when SLCL was possible (Condition R). This pattern clearly supports the Prioritize-and-Complement Hypothesis, suggesting that CSWL was not employed because SLCL was perfectly disambiguating.

Experiment 3, finally, evaluated the interaction of SLCL and CSWL when both are in conflict. As in Experiments 1 and 2, learning was above chance, strengthening the validity of the Multiple-Mechanisms Hypothesis. Importantly, as predicted by the Prioritize-and-Complement Hypothesis, we found that SLCL and CSWL competed to determine learning: Since SLCL was not perfectly disambiguating in any single trial, learners employed CSWL to complement their knowledge. However, both supported two different word meanings. Therefore a scenario of competition was created and findings suggest that the competing target meanings were learned approximately equally.

Experiments 1 to 3 therefore support both the Multiple-Mechanisms Hypothesis and the Prioritize-and-Complement Hypothesis. Concretely, the data reflects a priority for SLCL: If it is perfectly disambiguating, CSWL is not used (Condition R of Exp. 2) but if SLCL does not provide sufficient information, CSWL is additionally employed (Condition R2 of Exp.1). While results from Experiment 3 do not allow to conclude about the relative priority of either mechanism, it is likely, in the light of Experiment 2, that CSWL was only used because SLCL was not perfectly disambiguating. We would expect SLCL to block CSWL in a similar way as in Experiment 2 in a scenario in which SLCL and CSWL are in conflict and SLCL is perfectly disambiguating.

We suggest that the interplay of CSWL and SLCL can be seen as representative of the way word learning mechanisms can interact. Reliable top-down cues such as pointing may, for instance, have a similar interplay with CSWL as does SLCL. It is an interesting question how eye gaze may interact with both CSWL and SLCL: While in Nappa et al.'s (2009) study syntax overrode gaze, gaze may be stronger than CSWL. The interaction between verbal constraints and gaze may be similar as in Nappa's experiments, resulting in a priority for SLCL. While further research is needed to answer these questions, studies such as Nappa's and the one presented in this paper do make clear that because the use of a single mechanism is in fact influenced by the co-availability of other cues and mechanisms, investigating scenarios of interplay is essential for defining each mechanism's role and, therefore, the word learner's actual behavior.

We also found support for the hypothesis that CSWL is (or at least can be) a probabilistic and parallel learning mechanism: Learners in Condition N of Experiment 2 were able to differentiate between small differences in co-occurrence frequencies and kept track of more than just the currently best candidate. Multiple-hypothesis tracking is also supported by eye-movements in trials which were influenced by CSWL (Experiment 1: Condition N and R2; Experiment 2: Condition N): In contrast to trials in which exactly one referent was supported (i.e., when SLCL was possible), different candidates were looked at. This finding is in line with the data from Vouloumanos et al. (2008) but goes beyond their findings, revealing that CSWL can still operate in a parallel manner when learning trials are ambiguous (referential uncertainty) and situated (nouns embedded in sentences and objects embedded in scenes). We acknowledge that these conclusions may not generalize to every learning scenario. Learners may be capable of multiple-hypothesis tracking only to a certain level of complexity of the input and learning scenario and otherwise fail in this task, for instance, when too many possible referents are co-present with a novel noun. Indeed our results even support the hypothesis that CSWL is a learning mechanism which is not always influencing the learner: When more deterministic cues are available, cross-situational statistics may be ignored. However, that does not necessarily imply that CSWL-friendly situations do not exist and affect word learning at least to a certain extent.

Regarding the nature and role of sentence-level constraints, we found that they have *the potential* to offer a deterministic mechanism for word learning, as eye-movements in Experiments 1 and 2 reveal: Participants rapidly (even predictively) attended to the depicted objects which matched the verbal restrictions. We suggest that it is due to this highly disambiguating cue that SLCL also led to better learning than CSWL alone. Sentence-level constraints are clearly not always as disambiguating as in our word-learning experiments. Moreover, it certainly is possible that verbal constraints have a stronger influence in our experiments than in natural situations since learners are implicitly made aware of the relevance of the verb's meanings by starting the experiment with a verb-learning phase. However, it seems plausible to suggest that sentence-level constraints do often have a highly disambiguating potential because they build on prior linguistic and world knowledge which any adult is an expert in. It is this prior knowledge abut linguistic regularities and states and relations in the world which makes verbal constraints a generally highly reliable cue.

Another remark concerns the transferability of these findings and conclusions to child word learning. It seems plausible to suggest that infants, to the extent that they are old enough to have knowledge about sentence structures (i.e. from about 2 years of age), could use sentence-level constraints and cross-situational evidence simultaneously. Various syntactic bootstrapping studies reveal that 2-3 year old children are are able to combine observational cues and linguistic information to make inferences about word meanings (e.g., Lee & Naigles, 2008). However, multiple-cue integration capabilities are subject to development (e.g., Trueswell, Sekerina, Hill, & Logrip, 1999; Snedeker and Trueswell, 2004), which means that, depending on the exact age of a child, more cues do not have to result in better learning. As a study by Lidz, Bunger, Leddon, Baier, & Waxman (2010) reveals, 22-month-old infants performed better in verb learning when visual information (videos) was accompanied by sentences which were missing lexical information about the subject (i.e., when the subject was a pronoun: *It is blicking*) than when the lexical information was provided (*The flower is blicking*). It is therefore an open question which children (i.e., children of which age) could benefit from combining which word learning mechanisms. With regard to the influence of cues in conflict, (Nappa et al., 2009) demonstrates that children's (3, 4, and 5 years old) as well as adult's reliance on linguistic information can be high and in fact even higher than the use of gaze cues, if both contradict each other. While this reveals that children (as adults) may have a bias to use linguistic information in certain situations, it is, as well, an open question what this exactly means regarding the interplay of CSWL and SLCL. Further research is needed for evaluation.

Finally, it is important to note that there is no doubt that the way learning operates cannot be precisely predicted for any learner and any situation. Different learners likely have different priorities and this is most likely the case for different learner groups (e.g., children versus adults, native speakers versus second language learners). Moreover, very situation-dependent factors such as attention may play a role for the learner's behavior and strategy. Priorities in a learner may also change temporarily or permanently, for instance in case a cue is learned to be more or less reliable than originally assumed. Our results therefore only reflect general biases and behavioral probabilities. More research is needed to systematically examine individual and further situational differences. However, the presented data suggest that learners prioritize mechanisms that are more deterministic, and invoke complementary mechanisms only when necessary. This, in turn, may result in competition.

6 Summary

In this paper, we evaluated the relative importance of two word-learning mechanisms: cross-situational word learning (CSWL) and word learning based on sentence-level constraints (SLCL). Both mechanisms are fundamentally different in nature: CSWL is a bottom-up way of learning which is based on cues that are relatively unreliable and resource intensive. SLCL, on the contrary, is a top-down mechanism and operates on potentially highly reliable and disambiguating cues. We examined whether and how both mechanisms are employed in three different scenarios of interaction, that is, as complementary, redundantly applicable, and conflicting. Results from three experiments (each employing one of these scenarios) support the hypothesis that learners are capable of dealing with more than one applicable word learning mechanism simultaneously. Moreover, findings from all three experiments support the hypothesis that learners prioritize the mechanism they deem most reliable and only consider an additional one in case the prioritized mechanism does not provide perfectly disambiguating cues. If the second mechanism supports the cues given by the first, it simply complements the learner's knowledge. If, however, the second mechanism suggests a contradicting solution (i.e., another word meaning), a situation of conflict arises.

7 Acknowledgments

We thank Padraic Monaghan and two anonymous reviewers for their helpful comments which greatly helped to improve this paper.

References

Akhtar, N. & Montague, L. (1999). Early lexical acquisition: The role of cross-situational learning. *First Language*, 19(57), 347–358.

Altmann, G. T. M. & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247–64.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.

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

Breitenstein, C. & Knecht, S. (2003). Language acquisition and statistical learning. *Der Nervenarzt*, 74(2), 133–43.

Carey, S. (1978). The child as a word learner. In J. Bresnan & G. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Childers, J. B. & Paik, J. H. (2008). Korean- and English-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language*, 36(1), 201–24.

Fernald, A., Zangl, R., Luz Portillo, A., & Marchman, V. (2008). Looking while listening. In I. Sekerina, E. Fernandez, & H. Clahsen (Eds.), *Developmental Psycholinguistics. On-line methods in children's language processing* (pp. 97–135). Amsterdam/Philadelphia: John Benjamins.

Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down? *System*, 32(3), 363–377.

Fisher, C. (2002). Structural limits on verb mapping: The role of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science*, 5(1), 55–64.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73(2), 135–76.

Gleitman, L., Cassidy, K., Nappa, R., Papagragou, A., & Trueswell, J. (2005). Hard words. *Language Learning and Development*, 1(1), 23–64.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133–156.

Knoeferle, P. & Crocker, M. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, 57(4), 519–543.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: evidence from eye-movements in depicted events. *Cognition*, 95(1), 95–127.

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

Lidz, J., Bunger, A., Leddon, E., Baier, R., & Waxman, S. R. (2010). When one cue is better than two: Lexical vs. syntactic cues to verb learning. Unpublished manuscript.

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

Medina, T., Snedeker, J., Trueswell, J., & Gleitman, L. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 914–919.

Monaghan, P. & Mattock, K. (2009). Cross-situational language learning: The effects of grammatical categories as constraints on referential labeling. In N. Taatgen, J. Van Rijn, J. Nerbonne,

& L. Schomaker (Eds.), *Proceedings of the 31st Annual meeting of the Cognitive Science Society* (pp. 2226–2231). Austin, TX: Cognitive Science Society.

Monaghan, P. & Mattock, K. (2012). Integrating constraints for learning word-referent mappings. *Cognition*, 123, 133–143.

Nappa, R., Wessel, A., McEldoon, K., Gleitman, L., & Trueswell, J. (2009). Use of speaker's gaze and syntax in verb learning. *Language Learning and Development*, 5(4), 203–234.

Quine, W. (1960). *Word and Object*. Cambridge, MA: Cambridge University Press.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61(1-2), 39–91.

Smith, K., Smith, A. D. M., & Blythe, R. A. (2010). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, (pp. 1–19).

Smith, L. & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106(3), 1558–68.

Snedeker, J. & Gleitman, L. (2004). Why is it hard to label our concepts. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 257–295). Cambridge: MIT Press.

Snedeker, J. & Trueswell, J. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, 49(3), 238–299.

Tannenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science*, 268(5217), 1632–1634.

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

Trueswell, J., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73, 89–134.

Vogt, P. & Smith, A. (2005). Learning colour words is slow: A cross-situational learning account. *Behavioral and Brain Sciences*, 28, 509–510.

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, 107(2), 729–42.

Vouloumanos, A. & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, 45(6), 1611–7.

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2007). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, 56(3), 165–209.

Yu, C. & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science: A Journal of the American Psychological Society*, 18(5), 414–20.

Yurovsky, D., Yu, C., & Smith, L. (2013). Competitive processes in cross-situational word learning. *Cognitive Science*, 37, 891–921.

Tables

Tab. 1: Noun learning percentages (t-tests against chance = 25%) Exp. 1

Verb Restriction	Blocks 1-2	Block 1	Block 2
all	**72%**$(t(62) = 12.18, p < .001)$	**62%**$(t(62) = 6.90, p < .001)$	**83%**$(t(62) = 14.24, p < .001)$
R1	**77%**$(t(62) = 10.04, p < .001)$	**69%**$(t(62) = 6.24, p < .001)$	**85%**$(t(62) = 12.56, p < .001)$
R2	**74%**$(t(62) = 9.25, p < .001)$	**64%**$(t(62) = 5.19, p < .001)$	**85%**$(t(62) = 11.34, p < .001)$
N	**66%**$(t(62) = 7.43, p < .001)$	**52%**$(t(62) = 3.49, p < .001)$	**80%**$(t(62) = 8.69, p < .001)$

Tab. 2: Confidence ratings Exp. 1

	Verb Restriction	Blocks 1-2	Block 1	Block 2
1	*all*	5.73	5.06	6.39
2	*R1*	6.98	6.34	7.50
3	*R2*	6.42	5.88	6.80
4	*N*	5.40	4.45	6.02

Tab. 3: Lmer models & p-values from MCMC sampling for confidence ratings Exp. 1, Blocks 1-2 [3]

	Predictor	Coef.	SE	t	MCMCmean	pMCMC	$Pr(>\|t\|)$
1	(Int) (R1)	7.063	0.379	18.648	7.041	.000	$< .001$
2	R2	-0.649	0.285	-2.274	-0.649	.038	$< .050$
3	N	-1.758	0.301	-5.842	-1.706	.000	$< .001$
4	(Int) (R2)	6.415	0.381	16.831	6.416	.000	$< .001$
5	R1	0.649	0.285	2.274	.0625	.039	$< .050$
6	N	-1.109	0.299	-3.701	-1.084	.001	$< .001$

[3] $ConfidenceRating \sim VerbRestriction + (1|sub) + (1|item)$

[4] $InspectionsduringNP2 \sim AOI + (1|sub) + (1|item) + (1|list), family = binomial(link = "logit")$

[5] Numbering of distractors here and elsewhere random

[6] $Inspectionsduringverb \sim AOI + (1|sub) + (1|item) + (1|list), family = binomial(link = "logit")$

Tab. 4: Lmer models for inspections on target vs. distractors & Distractor1/competitor vs. other AOIs during NP2 in Exp. 1, Blocks 1-2
tar = target, char = character, com = competitor, di = distractor [4] [5]

	Predictor	Coef.	SE	Wald z	p
R1					
1	(Int) (tar)	−0.116	0.136	−0.854	= .393
2	char	−1.342	0.169	−7.959	< .001
3	di1	−0.698	0.167	−4.172	< .001
4	di2	−0.746	0.168	−4.430	< .001
5	di3	−0.460	0.170	−2.707	< .010
6	(Int) (di1)	−0.814	0.143	−5.711	< .001
7	tar	0.698	0.167	4.172	< .001
8	char	−0.644	0.174	−3.702	< .001
9	di2	−0.047	0.174	−0.273	= .785
10	di3	0.238	0.175	1.359	= .174
11	(Int) (char)	−1.459	0.143	−10.186	< .001
12	tar	1.342	0.169	7.959	< .001
13	di1	0.644	0.174	3.702	< .001
14	di2	0.597	0.175	3.409	< .001
15	di3	0.882	0.177	4.999	< .001
R2					
16	(Int) (tar)	0.157	0.121	1.294	= .196
17	char	−1.381	0.164	−8.407	< .001
18	com	−0.505	0.159	−3.169	< .010
19	di2	−1.449	0.180	−8.046	< .001
20	di3	−1.308	0.182	−7.194	< .001
21	(Int) (com)	−0.348	0.128	−2.711	< .010
22	tar	0.505	0.159	3.169	< .010
23	char	−0.876	0.169	−5.181	< .001
24	di2	−0.944	0.184	−5.128	< .001
25	di3	−0.803	0.186	−4.316	< .001
26	(Int) (char)	−1.219	0.128	−9.530	< .001
27	tar	1.380	0.164	8.411	< .001
28	com	0.882	0.169	5.221	< .001
29	di2	−0.062	0.189	−0.331	= .741
30	di3	0.080	0.190	0.420	= .674
N					
31	(Int) (tar)	0.048	0.135	0.352	= .725
32	char	−1.743	0.175	−9.963	< .001
33	di1	−0.208	0.154	−1.350	= .177
34	di2	−0.660	0.163	−4.063	< .001
35	di3	−1.046	0.171	−6.102	< .001
36	(Int) (di1)	−0.161	0.135	−1.189	= .234
37	tar	0.208	0.1542	1.350	= .177
38	char	−1.535	0.175	−8.775	< .001
39	di2	−0.452	0.163	−2.778	< .010
40	di3	−0.838	0.172	−4.866	< .001
41	(Int) (char)	−1.691	0.148	−11.400	< .001
42	tar	1.750	0.175	10.010	< .001
43	di1	1.535	0.175	8.785	< .001
44	di2	1.096	0.182	6.008	< .001
45	di3	0.710	0.190	3.727	< .001

Tab. 5: Lmer models for inspections on target vs. distractors & Distractor1/competitor vs. other AOIs during the verb in Exp. 1, Blocks 1-2
tar = target, char = character, com = competitor, di = distractor [6]

	Predictor	Coef.	SE	Wald z	p
R1					
1	(Int) (tar)	0.016	0.158	0.104	$= .917$
2	char	-0.604	0.153	-3.937	$< .001$
3	di1	-0.634	0.164	-3.874	$< .001$
4	di2	-0.395	0.161	-2.450	$< .050$
5	di3	-0.337	0.167	-2.024	$< .050$
6	(Int) (di1)	-0.618	0.159	-3.883	$< .001$
7	tar	0.634	0.164	3.874	$< .001$
8	char	0.030	0.156	0.191	$= .848$
9	di2	0.239	0.164	1.458	$= .145$
10	di3	0.296	0.169	1.750	$= .080$
R2					
11	(Int) (tar)	-0.220	0.153	-1.439	$= .151$
12	char	-0.412	0.154	-2.670	$< .010$
13	com	0.142	0.160	0.889	$= .374$
14	di2	-0.161	0.163	-0.988	$= .323$
15	di3	-0.274	0.169	-1.642	$= .104$
16	(Int) (com)	-0.078	0.158	-0.495	$= .621$
17	tar	-0.142	0.160	-0.889	$= .374$
18	char	-0.554	0.158	-3.496	$< .001$
19	di2	-0.303	0.167	-1.820	$= .069$
20	di3	-0.416	0.172	-2.416	$< .050$
N					
21	(Int) (tar)	-0.021	0.147	-0.145	$= .885$
22	char	-0.897	0.155	-5.789	$< .001$
23	di1	-0.397	0.155	-2.556	$< .050$
24	di2	-0.237	0.159	-1.485	$= .138$
25	di3	-0.420	0.163	-2.580	$< .010$
26	(Int) (di1)	-0.419	0.148	-2.820	$< .010$
27	tar	0.397	0.155	2.556	$< .050$
28	char	-0.500	0.156	-3.206	$< .010$
29	di2	0.160	0.161	0.994	$= .320$
30	di3	-0.023	0.165	-0.142	$= .887$

Tab. 6: Example trials Exp. 2
Trials 1-3: Category 1, Trials 4-5: Category 2, Trials 6: Category 3

Trial	Sentence	Objects (co-occurrence frequencies in %)
Non-restrictive		
1	*Si laki gumbumema si daram* 'The man will point at the...'	corn (83), socks (50), gloves (17), hat (17)
2	*Si gadis tambamema si daram* 'The woman will take the...'	corn (83), socks (50), jumper (17), scarf (17)
3	*Si gadis gumbumema si daram* 'The woman will point at the...'	corn (83), socks (50), top (17), skirt (17)
4	*Si laki tambamema si daram* 'The man will take the...'	corn (83), cap (17), jacket (17), dress (17)
5	*Si laki gumbumema si daram* 'The man will point at the...'	corn (83), vest (17), jeans (17), coat (17)
6	*Si gadis tambamema si daram* 'The woman will point at the...'	burger (17), apple (17), shorts (17), apron (17)
Restrictive		
1	*Si laki bermamema si firel* 'The man will eat the...'	burger (83), jumper (50), gloves (17), hat (17)
2	*Si gadis bermamema si firel* 'The woman will eat the...'	burger (83), jumper (50), socks (17), scarf (17)
3	*Si gadis bermamema si firel* 'The woman will eat the...'	burger (83), jumper (50), top (17), skirt (17)
4	*Si laki bermamema si firel* 'The man will eat the...'	burger (83), cap (17), jacket (17), dress (17)
5	*Si laki bermamema si firel* 'The man will eat the...'	burger (83), vest (17), jeans (17), coat (17)
6	*Si gadis tambamema si firel* 'The woman will point at the...'	corn (17), apple (17), shorts (17), apron (17)

Tab. 7: Example test trials Exp. 2
VR = Verb Restriction, TT = Test Type

VR	TT	Noun	Objects (& co-occurrence frequencies in %)
N	1	*si daram*	corn (83), socks (50), gloves (17), scarf (17)
	2	*si daram*	hamburger (17), socks (50), jumper (17), apron (17)
R	1	*si firel*	hamburger (83), jumper (50), gloves (17), scarf (17)
	2	*si firel*	apple (17), jumper (50), shorts (17), apron (17)

Tab. 8: Lmer models for chosen meanings in conditions (low-frequency vs. high-frequency object, high-frequency meaning choices, low-frequency meaning choices), Exp. 2, Test Type 1, [7]

	Predictor	Coef.	SE	Wald z	p
low vs. high frequency					
1	(Int) (N)	-1.170	0.191	-6.133	$< .001$
2	R	-1.340	0.346	-3.876	$< .001$
amount high-frequency choices					
3	(Int) (N)	0.465	0.188	2.474	$< .050$
4	R	1.312	0.254	5.160	$< .001$
amount low-frequency choices					
5	(Int) (N)	-1.460	0.185	-7.891	$< .001$
6	R	-1.151	0.342	-3.369	$< .001$

[7] $ChosenMeaning \sim VerbRestriction + (1|sub) + (1|item), family = binomial(link = "logit")$

Tab. 9: Lmer models for chosen meanings in conditions (low-frequency object/category associate, category-associate choices, low-frequency meaning choices), Exp. 2, Test Type 2 [8]

	Predictor	Coef.	SE	Wald z	p
low vs. high frequency					
1	(Int) (N)	-0.224	0.255	-0.877	$= .380$
2	R	-1.177	0.301	-3.917	$< .001$
amount cat-as. choices					
3	(Int) (N)	-0.543	0.264	-2.058	$< .050$
4	R	0.605	0.223	2.711	$< .010$
amount low-frequency choices					
5	(Int) (N)	-0.830	0.169	-4.923	$< .001$
6	R	-1.084	0.267	-4.050	$< .001$

[8] $ChosenMeaning \sim VerbRestriction + (1|sub) + (1|item), family = binomial(link = "logit")$

Tab. 10: Pairwise comparisons for ANOVAs by subject (Bonferroni adjustment), Exp. 2, Test Type 1

	chosen meaning	chosen meaning	Mean Difference	SE	p
Condition N					
1	83%	50%	.417	.067	$< .001$
2	83%	17%-1	.510	.053	$< .001$
3	83%	17%-2	.495	.070	$< .001$
4	50%	83%	-.417	.067	$< .001$
5	50%	17%-1	.094	.042	$= .215$
6	50%	17%-2	.078	.046	$= .603$
7	17%-1	83%	-.510	.053	$< .001$
8	17%-1	50%	-.094	.042	$= .215$
9	17%-1	17%-2	.016	.035	$= 1.000$
10	17%-2	83%	-.495	.070	$< .001$
11	17%-2	50%	-.078	.046	$= .603$
12	17%-2	17%-1	.016	.035	$= 1.000$
Condition R					
13	83%	50%	.766	.050	$< .001$
14	83%	17%-1	.786	.050	$< .001$
15	83%	17%-2	.792	.047	$< .001$
16	50%	83%	-.766	.050	$< .001$
17	50%	17%-1	.021	.018	$= 1.000$
18	50%	17%-2	.026	.020	$= 1.000$
19	17%-1	83%	-.786	.050	$< .001$
20	17%-1	50%	-.021	.018	$= 1.000$
21	17%-1	17%-2	.005	.016	$= 1.000$
22	17%-2	83%	-.792	.047	$< .001$
23	17%-2	50%	-.026	.020	$= 1.000$
24	17%-2	17%-1	-.005	.016	$= 1.000$

Tab. 11: Pairwise comparisons for ANOVAs by subject (Bonferroni adjustment), Exp. 2, Test Type 2; CA = category associate

	chosen	chosen	Mean Difference	SE	p
Condition N					
1	CA	50%	.083	.074	$= 1.00$
2	CA	17%-1	.266	.058	$< .010$
3	CA	17%-2	.214	.060	$< .050$
4	50%	83%	$-.083$	.074	$= 1.00$
5	50%	17%-1	.182	.045	$< .010$
6	50%	17%-2	.130	.042	$< .050$
7	17%-1	CA%	-.266	.058	$< .010$
8	17%-1	50%	-.182	.045	$< .010$
9	17%-1	17%-2	-.052	.039	$= 1.000$
10	17%-2	CA%	-.214	.060	$< .010$
11	17%-2	50%	-.130	.042	$< .050$
12	17%-2	17%-1	.052	.039	$= 1.000$
Condition R					
13	CA	50%	.375	.079	$< .010$
14	CA	17%-1	.375	.079	$< .010$
15	CA	17%-2	.286	.089	$< .050$
16	50%	83%	$-.375$	.079	$< .010$
17	50%	17%-1	.000	.032	$= 1.00$
18	50%	17%-2	$-.089$	.046	$= .402$
19	17%-1	CA%	-.375	.079	$< .001$
20	17%-1	50%	.000	.032	$= .215$
21	17%-1	17%-2	-.089	.037	$= 1.000$
22	17%-2	CA%	-.286	.089	$< .050$
23	17%-2	50%	.089	.046	$= .402$
24	17%-2	17%-1	.089	.037	$= .156$

Tab. 12: Lmer models for inspections on AOIs during verb and NP2, Exp. 2, Trial Category 1 [9]

		Predictor	Coef.	SE	Wald z	p
Condition N						
1	*V*	(Int) (83%)	0.298	0.125	2.378	$< .050$
2		50%	-0.056	0.129	-0.436	$= .663$
3		17%-1	-0.223	0.130	-1.711	$= .087$
4		17%-1	-0.283	0.130	-2.168	$< .050$
5	*NP2*	(Int) (83%)	0.050	0.105	0.481	$= .630$
6		50%	-0.276	0.127	-2.176	$< .050$
7		17%-1	-0.472	0.130	-3.640	$< .001$
8		17%-1	-0.225	0.129	-1.752	$= .080$
Condition R						
9	*V*	(Int) (83%)	1.066	0.123	8.639	$< .001$
10		50%	1.086	0.137	-7.937	$< .001$
11		17%-1	-1.233	0.139	-8.880	$< .001$
12		17%-1	-1.138	0.141	-8.085	$< .001$
13	*NP2*	(Int) (83%)	-0.336	0.124	-2.702	$< .010$
14		50%	-0.914	0.1418	-6.450	$< .001$
15		17-1	-1.167	0.151	-7.755	$< .001$
16		17%-1	-1.268	0.157	-8.103	$< .001$

[9] $Inspections \sim AOI + (1|sub) + (1|item) + (1|block), family = binomial(link = "logit")$

Tab. 13: Lmer models for looks to the 50% object during NP2, Exp. 2, Trial Category 1 [10]

	Predictor	Coef.	SE	Wald z	p
1	(Int.) (Cond. N)	0.034	0.120	0.281	$= .779$
2	Cond. R	-1.080	0.138	-7.822	$< .001$

[10] $Inspectionsto50\% \sim VerbRestriction + (1|sub) + (1|item), family = binomial(link = "logit"$

Tab. 14: Example trials Exp. 3

Category	Trial	Sentence	Objects (co-occurrence frequencies in %)
Condition N			
1	1	*Si laki gumbumema si daram* 'The man will take the DARAM.'	corn (83%), socks (50%), dress (17%)
	2	*Si gadis gumbumema si daram* 'The woman will take the DARAM.'	corn (83%), socks (50%), top (17%)
	3	*Si laki gumbumema si daram* 'The man will take the DARAM.'	corn (83%), socks (50%), pizza (17%)
2	4	*Si gadis tambamema si daram* 'The woman will take the DARAM.'	corn (83%), jacket (17%), jumper (17%)
	5	*Si gadis tambamema si daram* 'The woman will take the DARAM.'	corn (83%), skirt 17%)
3	6	*Si laki tambamema si daram* 'The man will take the DARAM.'	
Condition R			
1	1	*Si laki gumbumema si firel* 'The man will take the FIREL.'	dress (83%), apple (50%), socks (17%)
	2	*Si gadis gumbumema si firel* 'The woman will take the FIREL.'	dress (83%), apple (50%), top (17%)
	3	*Si laki gumbumema si firel* 'The man will take the FIREL.'	dress (83%), apple (50%), pizza (17%)
2	4	*Si gadis bermamema si firel* 'The woman will eat the FIREL.'	dress (83%), jacket (17%), jumper (17%)
	5	*Si gadis bermamema si firel* 'The woman will eat the FIREL.'	dress (83%), skirt (17%)
3	6	*Si laki bermamema si firel* 'The man will eat the FIREL.'	

Tab. 15: Lmer models for chosen meaning (low-frequency/high-frequency object), Exp. 3 [11]

	Predictor	Coef.	SE	Wald z	p
1	(Int) (Non-restrictive)	-3.5781	0.601	-5.950	$< .001$
2	Restrictive	3.425	0.532	6.442	$< .001$

[11] $ChosenMeaning \sim VerbRestriction + (1|sub) + (1|item), family = binomial(link = "logit")$

Figures

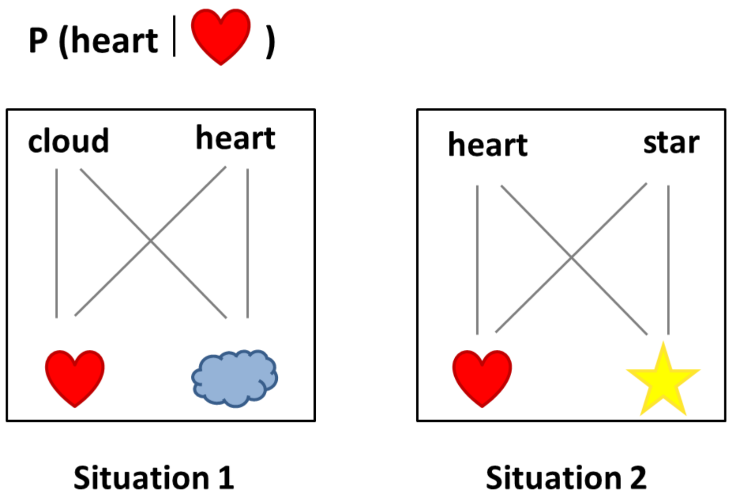

Fig. 1: Cross-situational word learning (example parallel to Smith & Yu, 2008: 1560)

Fig. 2: Example item, Exp. 1
R1: *Si gadis mautimema si sonis.* ('The woman will empty the bucket.')
R2: *Si gadis felimema si kemei.* ('The woman will iron the jeans.')
N: *Si gadis tambamema si worel.* ('The woman will take the broccoli.')

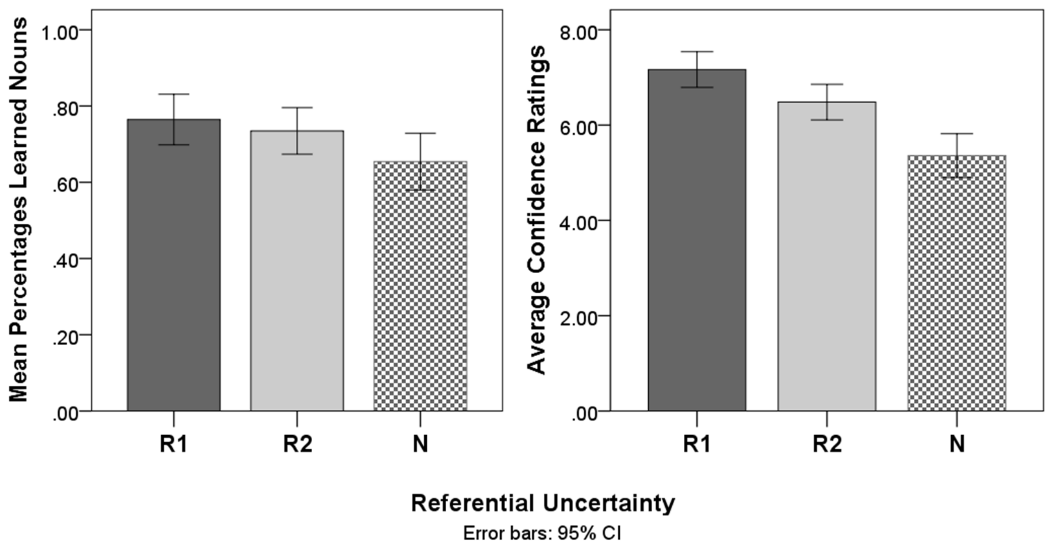

Fig. 3: Off-line data Exp. 1, Blocks 1-2: Mean percentages of learned nouns (left chart) and average confidence ratings (right chart)

[12]Proportions refer to proportions of trial with at least one inspection to AOI. Therefore bars do not add up to one.

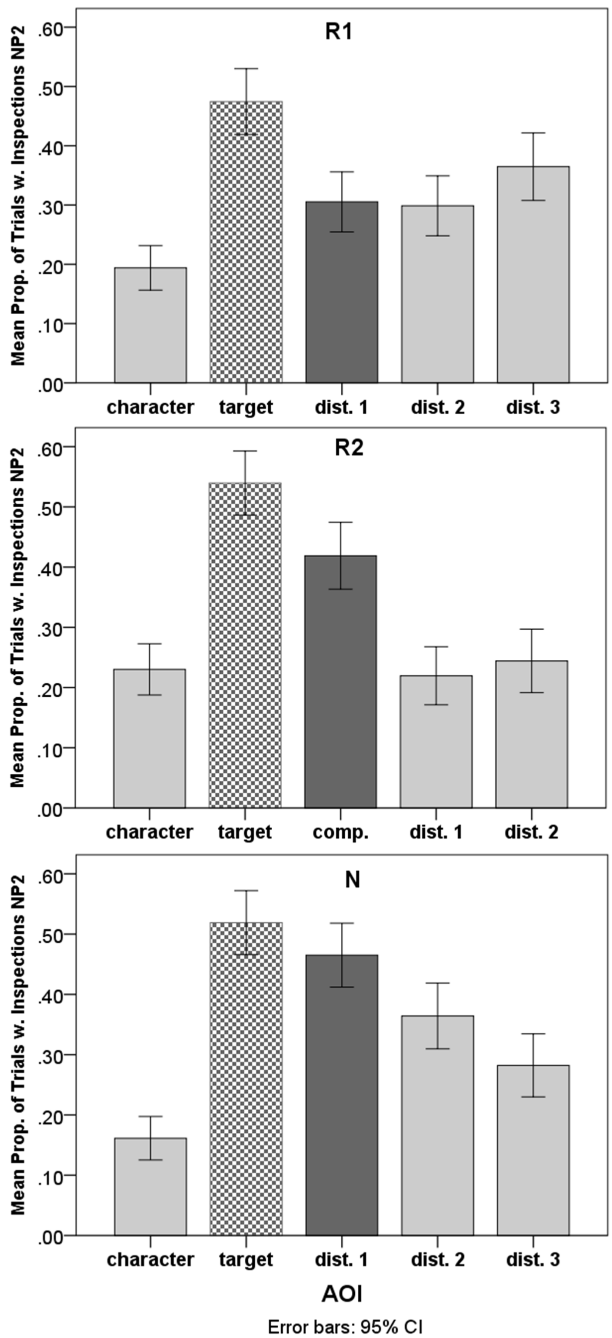

Fig. 4: Proportions of trials with at least one inspection during NP2 in Exp. 1, Blocks 1-2: Condition R1 (top), Condition R2 (mid), and Condition N (bottom) [12]

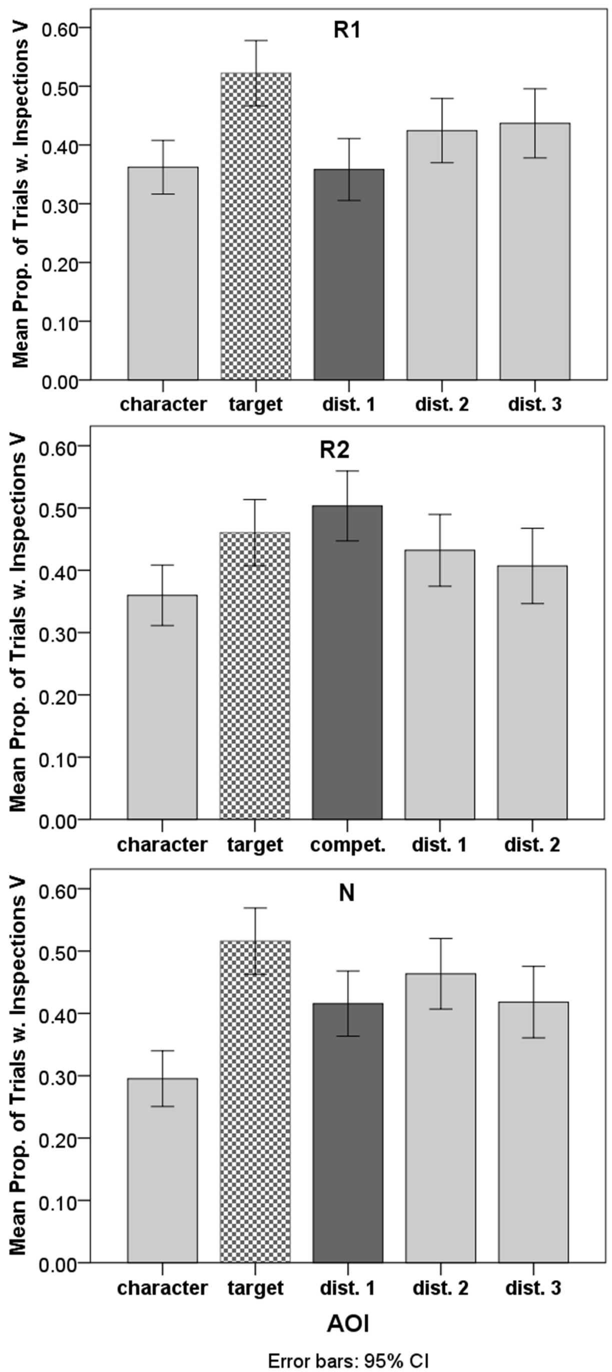

Fig. 5: Proportions of trials with at least one inspection to AOIs during the verb in Exp. 1, Blocks 1-2: Condition R1 (top chart), Condition R2 (mid chart), and Condition N (bottom chart)

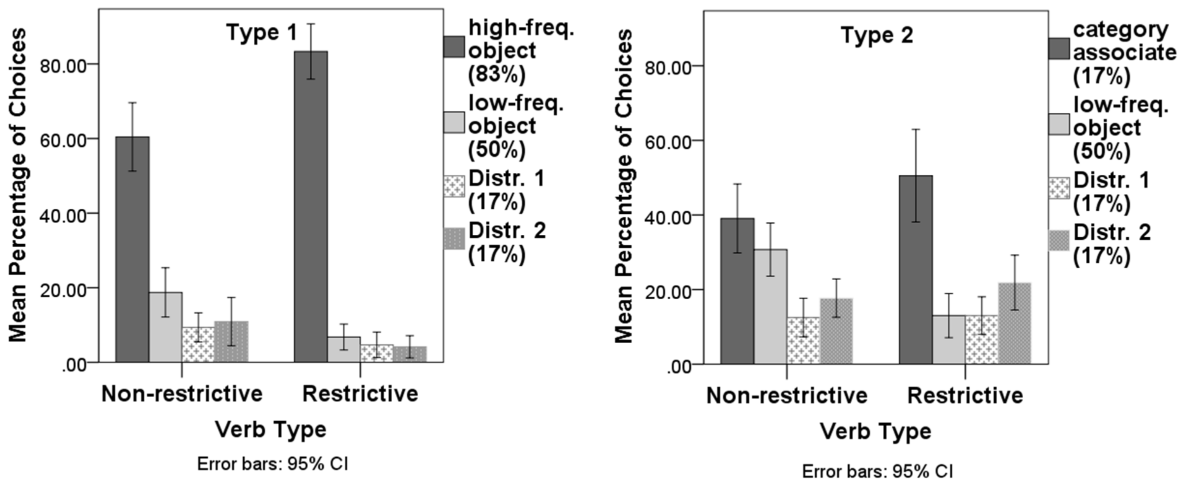

Fig. 6: Mean percentages of chosen meanings in vocabulary-test trials of Test Type 1 (left chart) and Test Type 2 (right chart), Exp. 2

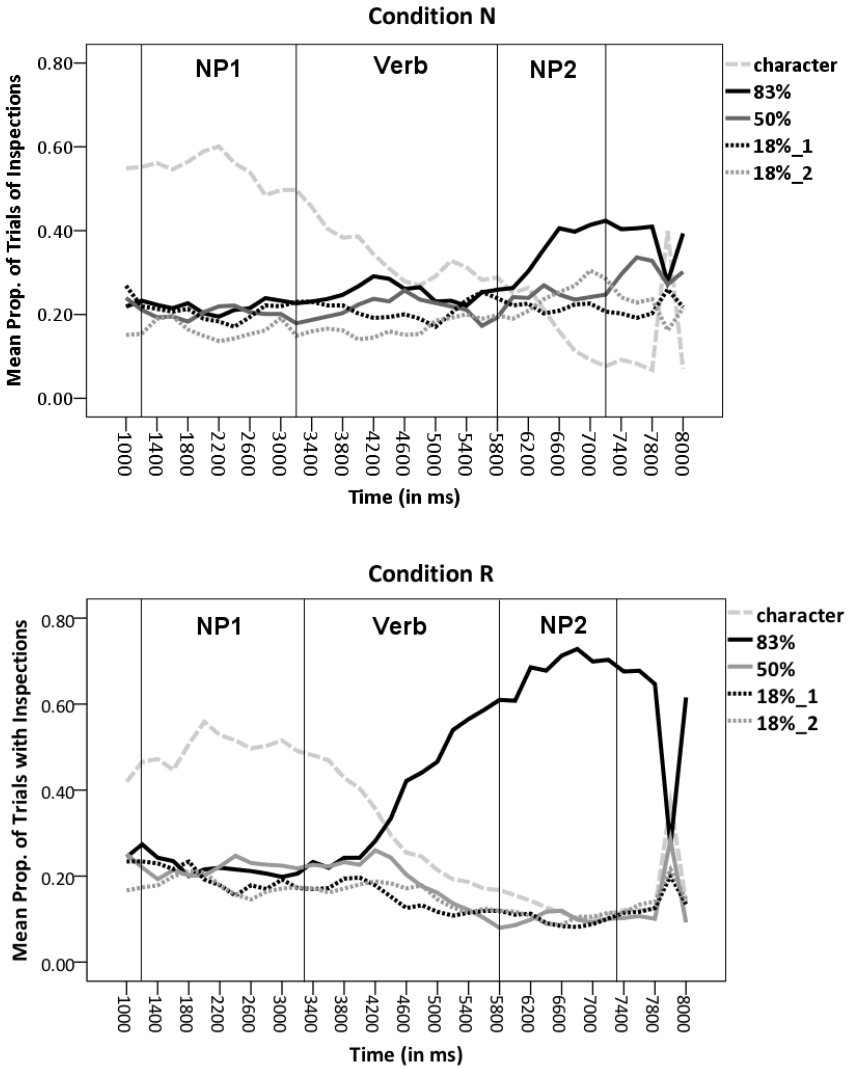

Fig. 7: Time graphs Exp. 2, Trial Category 1: Condition N (top) and Condition R (bottom) Mean proportions of trials with at least one inspection to AOIs (character, 83% object, 50% object, 17% objects) during NP1, verb, and NP2

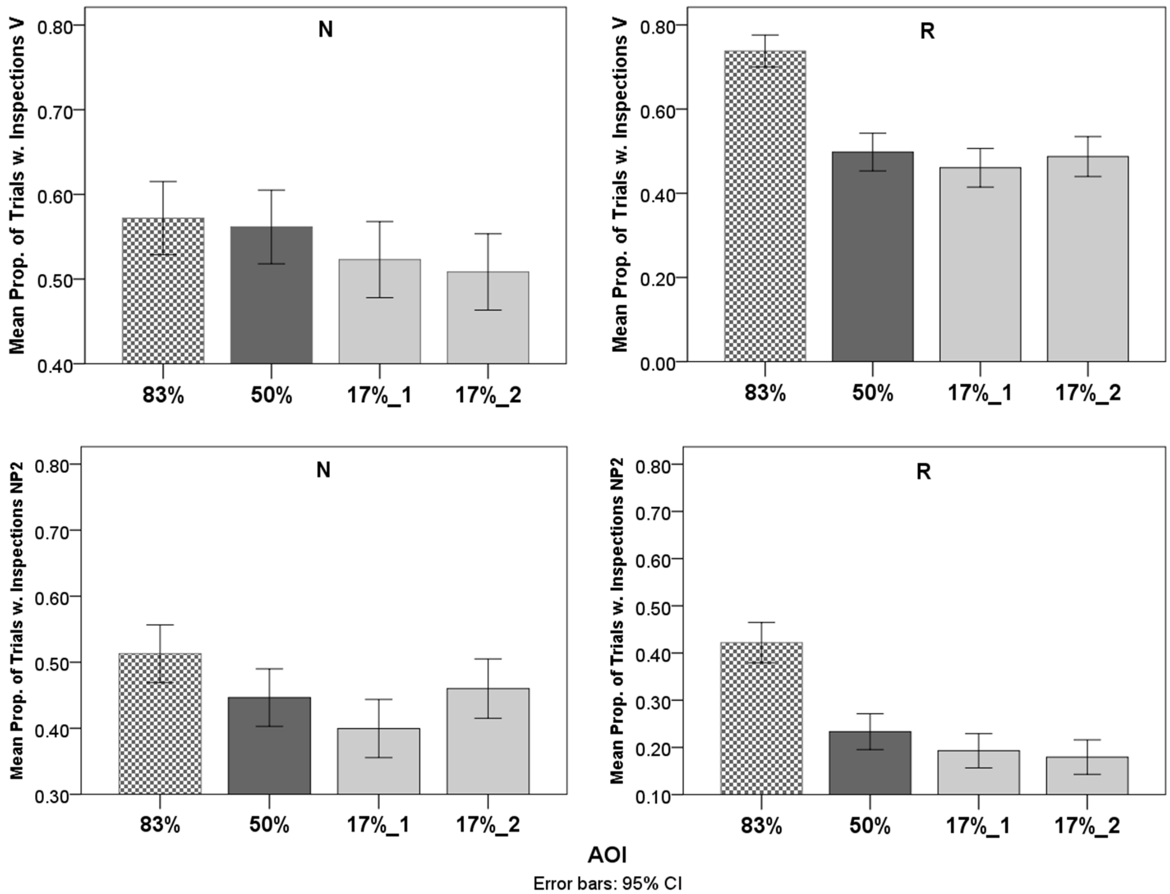

Fig. 8: Proportions of trials with at least one inspections to AOIs, Exp. 2, Trial Category 1: verb (top charts) and NP2 (bottom charts), Condition N (left charts) and Condition R (right charts)

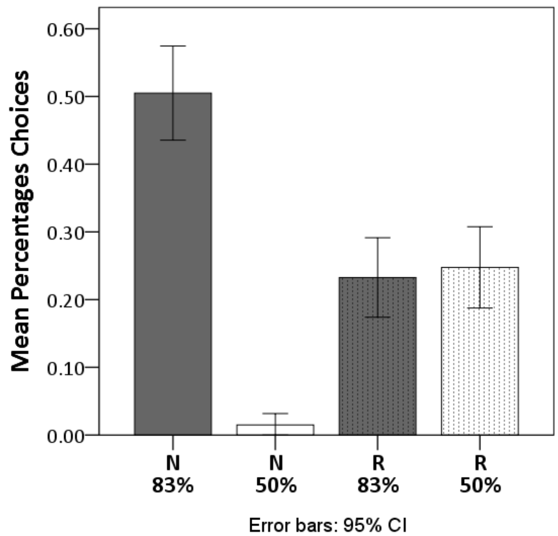

Fig. 9: Percentages of vocabulary-test choices (high-frequency object / low-frequency object) in conditions, Exp. 3