Integrating constraints for learning word-referent mappings

Padraic Monaghan & Karen Mattock

Centre for Research in Human Development and Learning, Department of Psychology

Lancaster University, UK

Abstract count: 150

Word count: 6196

Corresponding author:

Padraic Monaghan

Department of Psychology

Lancaster University

Lancaster

LA1 4YF

UK

E-mail: p.monaghan@lancaster.ac.uk

Tel: +44 1524 593813

Fax: +44 1524 593744

Abstract

Learning word-referent mappings is complex because the word and its referent tend to co-occur with multiple other words and potential referents. Such complexity has led to proposals for a host of constraints on learning, though how these constraints may interact has not yet been investigated in detail. In this paper, we investigated interactions between word co-occurrence constraints and cross-situational statistics in word learning. Analyses of child-directed speech revealed that when both object-referring and non-referring words occurred in the utterance, referring words were more likely to be preceded by a determiner than when the utterance contained only referring words. In a word learning study containing both referring and non-referring words, learning was facilitated when non-referring words contributed grammatical constraints analogous to determiners. The complexity of multi-word utterances provides an opportunity for co-occurrence constraints to contribute to word-referent mapping, and the learning mechanism is able to integrate these multiple sources of information.



*Key words*: word learning; cross-situational learning; child-directed speech; distributional information; multiple cue integration; corpus analysis

There are numerous sources of information present in the child's environment that potentially contribute to word-learning, such as representational constraints (Markman, 1990), social and gestural cues (Baldwin, 1993; Tomasello, 2003), cross-situational associations (Yu & Smith, 2007), between-word co-occurrences (Siskind, 1996; Yu & Ballard, 2007), and phonological and prosodic cues (Monaghan, Christiansen, & Chater, 2007). The operation of individual sources of information to constrain language learning has been a topic of considerable interest (Markman, 1990; White & Morgan, 2008), yet the integration of cues across different modalities has as yet received substantially less attention (though see notable exceptions by Gogate, Walker-Andrews, and Bahrick (2001), Hollich, Hirsch-Pasek, and Golinkoff (2000), and Monaghan et al. (2007)).

Each source of information increases the complexity of the learning situation, but also enhances the possibilities for multiple, interacting cues to constrain learning. The importance of this interacting information becomes more critical as environmental complexities are also considered in word learning. For instance, in the child's environment on hearing a particular word, there are likely to be multiple possible referents for that word present in the environment. However, over time, the child is likely to observe a co-occurrence of a particular word and a particular referent in the environment from multiple learning instances (Gleitman, 1990; Horst, Scott, & Pollard, 2010; Pinker, 1989; Smith, Smith, & Blythe, 2011). Yu and Smith (2007) illustrated a solution to this difficulty of multiple word-object correspondences in individual learning situations by demonstrating experimentally that learners can derive word-referent associations from multiple presentations. Participants were presented with several nonsense words and simultaneously with several novel objects on a computer screen.

From a single presentation, knowing which of the words referred to which object was impossible, but over multiple learning situations participants were able to exploit the co-occurrence of particular words with particular objects. This was because there was a correlation between occurrence of a particular word and the particular object to which it referred in the environment, whereas chance co-occurrences of words and objects reduced in reliability across increasing numbers of situations. Smith and Yu (2008) found that 12- and 14-month-old infants were also sensitive to similar cross-situational statistics for learning word-object mappings.

In these previous studies of cross-situational learning (Smith & Yu, 2008; Yu & Smith, 2007), the learning situation has provided a perfect correspondence for word-referent mappings to the learner in that all the words heard by the learner have referents, and all objects perceived are referred to in each situation. In these studies, the probability of a word occurring given its referent appearing was 1.0, and similarly, the probability of a referent appearing given its word being heard was also 1.0. Thus, participants could be exploiting pragmatic constraints in assuming that all words and objects provide information. In addition, learners could be applying mutual exclusivity constraints to learning (e.g., Markman, 1990; White & Morgan, 2008). Mutual exclusivity refers to the assumption by the learner that there is only one word used to refer to each referent, thus learning the name of one object in the array can assist in directing the learner to learn to attach a label to another object. Smith, Smith, and Blythe (2009) provide a useful discussion for how this applies to cross-situational learning in the paradigms used by Yu and Smith (2007).

Yet, the child's learning experience is still more complex, in that not all words within utterances have referents in the environment, and the child must learn which are the words in speech that potentially can refer to objects in the environment (Yu & Ballard, 2007), which we term "referring words". Given that the mean length of utterance in child-directed speech is in the range 3 to 4 (Monaghan & Christiansen, 2010) this suggests that in learning to attach a label to an object the child has to disregard an average of two words as the potential referring word. This additional requirement to disregard potential referring words has not yet been tested in previous studies of cross-situational learning (Horst et al., 2010; Smith et al., 2011; Smith & Yu, 2008; Yu & Smith, 2007). It also entails that constraints such as mutual exclusivity cannot reliably provide information about currently unknown word-referent mappings as the speech may contain either referring words or words that cannot refer to objects in the environment (such as verbs, adverbs, or most function words), which we refer to as "non-referring words"., Consequently, knowing the name for one of the objects in the environment will not assist in distinguishing between such referring and non-referring words in speech.

Yu and Ballard (2007) tested a computational model that learned associations between words and objects in the child's environment in small corpora of child-directed speech, so taking into account the additional complexity of natural language utterances. The model was successful in learning associations between words and objects, and was also successful in learning that certain non-referring words were not associated with objects, despite frequently co-occurring with them. Similarly, the model was generally successful in learning that function words were not associated with particular objects, despite their high frequency of co-occurrence. Yu (2006) tested a similar computational model on a small corpus of child-directed speech, but also added syntactic information about each word in terms of acquired “equivalence classes” of words. He found that learning associations between words and objects was assisted by the presence of syntactic information – the words in the classes corresponding to concrete nouns received a boost in associative strength within his model, increasing their salience as potential referring words.

Yet, Yu and Ballard’s (2007) and Yu’s (2006) studies are necessarily limited to a small corpus of child-directed speech, because it required hand-coding the child’s environment. Our first major aim was to examine the nature of the language environment in larger corpora of child-directed speech to test the extent to which referring and non-referring words occurred in speech. In addition, we also aimed to assess the possibility that non-referring words in child-directed speech may provide additional constraints that potentially assist in learning the mapping between referring words and referents. Distributional analyses of child-directed speech have shown that sentence context provides valuable information about grammatical category (Mintz, Newport, & Bever, 2002; Monaghan et al., 2007; St Clair, Monaghan, & Christiansen, 2010), and Yu (2006) demonstrated how an associative learner may be able to exploit this information for determining referring words in speech. The analyses we present also tested whether the reliability of this distributional information varies according to the presence or absence of other non-referring words. We predicted that when the ambiguity of referring words increased in the utterance then contextual information would be more reliable in child-directed speech. If this is the case, then it suggests either that the speaker dynamically adjusts the contextual information present within their utterances to assist in acquisition of word-referent mappings, or that a side-effect of complexity is greater information for reference for the language learner. This was tested in the first study.

The second major aim was to determine whether cross-situational learning is robust in the presence of non-referring words in speech. The computational study of Yu and Ballard (2007) suggests that the correspondences between words in child directed speech and objects in the environment can be learned robustly, and that non-referring and function words (words that typically indicate a syntactic relationship in an utterance, rather than have an independent meaning) do not chronically disturb this process. The second study provides an empirical test of the robustness of this learning.

The additional complexity of non-referring words provides a challenge to the learner, but it also creates an opportunity for additional language-internal constraints to assist in learning word-referent mappings. Learning the grammar of the language constrains which words may potentially relate to referents in the environment. Hence, the third major aim, again tested in study 2, was to investigate the role of function words in the utterance as additional constraints on cross-situational learning. For instance, learning that determiners “the” or “a” typically precede nouns enables these function words to be ignored as potential referring words themselves and provides a focus on the noun as the reference word. Children are able to distinguish function and content words as young as 10 months (Shady, Gerken, & Jusczyk, 1995), by 14 months infants are able to use determiners to identify nouns (Höhle, Weissenborn, Kiefer, Schulz, & Schmitz, 2004), and by 17 months determiners assist in distinguishing nouns from proper nouns (Katz, Baker, & MacNamara, 1974), indicating that function words can support mapping from nouns to objects early in the child’s language development.

Table 1. Properties of the child-directed speech corpora.

| Corpus | Number of Utterances | Number of Words | Mean Utterance Length |
|---|---|---|---|
| Anne | 26,224 | 95,956 | 3.66 |
| Aran | 20,867 | 106,998 | 5.13 |
| Eve | 16,964 | 58,521 | 3.45 |
| Naomi | 9,128 | 29,314 | 3.19 |
| Nina | 17,011 | 73,411 | 4.32 |
| Peter | 21,008 | 74,599 | 3.55 |

## Study 1: Corpus Analyses of Child-Directed Speech

### *Method*

*Corpus preparation.* We analysed child-directed speech addressed to six children under the age of 2;6, taken from the CHILDES database (MacWhinney, 2000). The corpora were: Anne and Aran (Theakston, Lieven, Pine, & Rowland, 2001), Eve (Brown, 1973), Naomi (Sachs, 1983), Nina (Suppes, 1974), and Peter (Bloom, Hood, & Lightbown, 1974). Each corpus was automatically tagged by a parser with 95% accuracy (Sagae, MacWhinney, & Lavie, 2004). Properties of each corpus are shown in Table 1.

*Corpus analysis*. Each utterance was assessed in terms of the grammatical categories of words it contained. As we were interested in the co-occurrence of words with objects, we selected only those utterances that contained at least one noun, thereby extending Yu and Ballard's (2007) analyses to larger corpora.

For utterances containing at least one noun, we counted the number of words from other grammatical categories they contained, that could not be paired with referents. In particular, we measured other content words (words which carry meaning rather than just a syntactic role). Certain verbs may be used in very specific situations, and so could provide misleading information to the child about the identity of the referring word in speech, such as the verb "watch" that occurs in the same utterance as more than 20% of occurrences of "television" in the whole CHILDES corpus. Similarly, adjectives and adverbs could be misleading in terms of their link to potential referents, with over 30% of occurrences of "hairy" preceding "monster" in the CHILDES corpus. Though the corpus analyses provide only indirect evidence for co-occurences of words other than nouns with particular objects, we know from the work of Yu and Ballard (2007) that such co-occurrences are commonplace in the child's learning environment.

We also measured occurrence of function words, comprising determiners, conjunctions, and prepositions, that are likely to occur with a referring noun, but, unlike content words, occur frequently in speech. The child has to learn that the lack of variation in these words' usage indicates that they are poor candidates for words mapping to objects as they co-occur with nouns and the objects to which the noun refers, but are not specific to any particular referring word. Table 4 in Yu and Ballard (2007), for instance, indicates that "a" and the object "bird" co-occur in the child's environment almost as often as the word "bird" does (18 versus 22 co-occurrences, respectively).

As an example of this in the corpora we analyse in Study 1, consider the following utterances from the Anne corpus.

Mother*: Do you think the penguin wants a sausage?*

Mother*: Chocolate*

Mother*: Is the penguin eating the chocolate?*

The first utterance contains two nouns, each preceded by a determiner (*the penguin, a sausage*), so in this case the nouns are marked by function words indicating their grammatical role. If the child is sensitive to the distributional information from these frequent words, *a* and *the*, then this will help to constrain the possible words that the child can use to refer to objects in their environment. The second utterance contains a single word – a noun. In this case, there is only a single word that could map to a potential object in the environment. The third utterance is also an instance where both nouns are preceded by determiners (*the penguin, the chocolate*).

Aslin, Woodward, LaMendola, and Bever (1996) instructed parents to teach 12-month-old infants a novel word, and they tended to use the word in multi-word utterances including verbs and function words, suggesting that the general pattern of utterances for learning word-object pairings is not qualitatively distinct from general patterns of parent-child discourse in terms of the range of grammatical items used.

*Results and discussion*

Figure 1 shows the number of utterances containing different numbers of nouns. Whereas the learning situation of several nouns occurring in an utterance is not infrequent, the typical exposure the child experiences is of utterances containing either just one word as

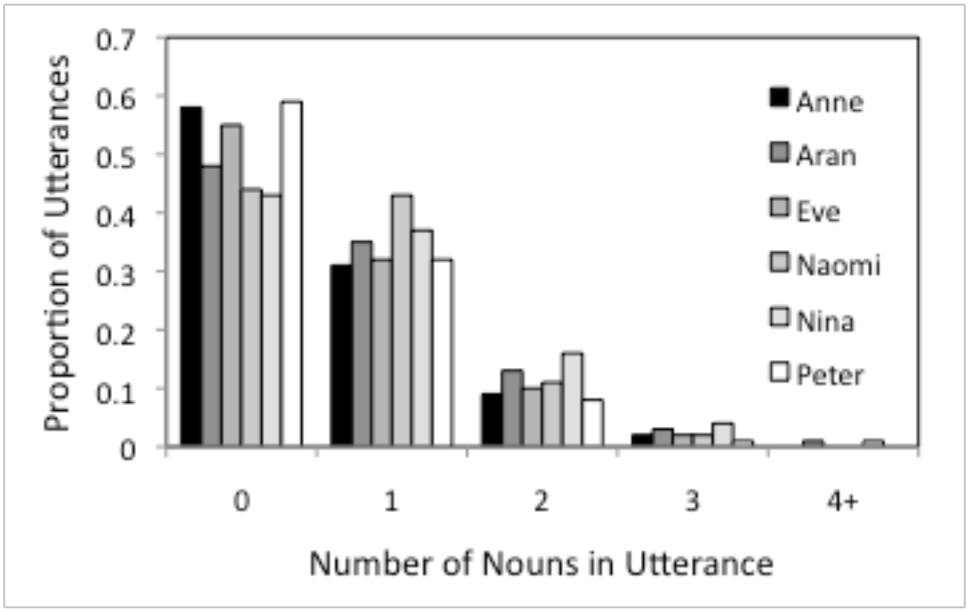


Figure 1. Proportion of utterances containing different numbers of nouns for each of the 6 child-directed speech corpora.

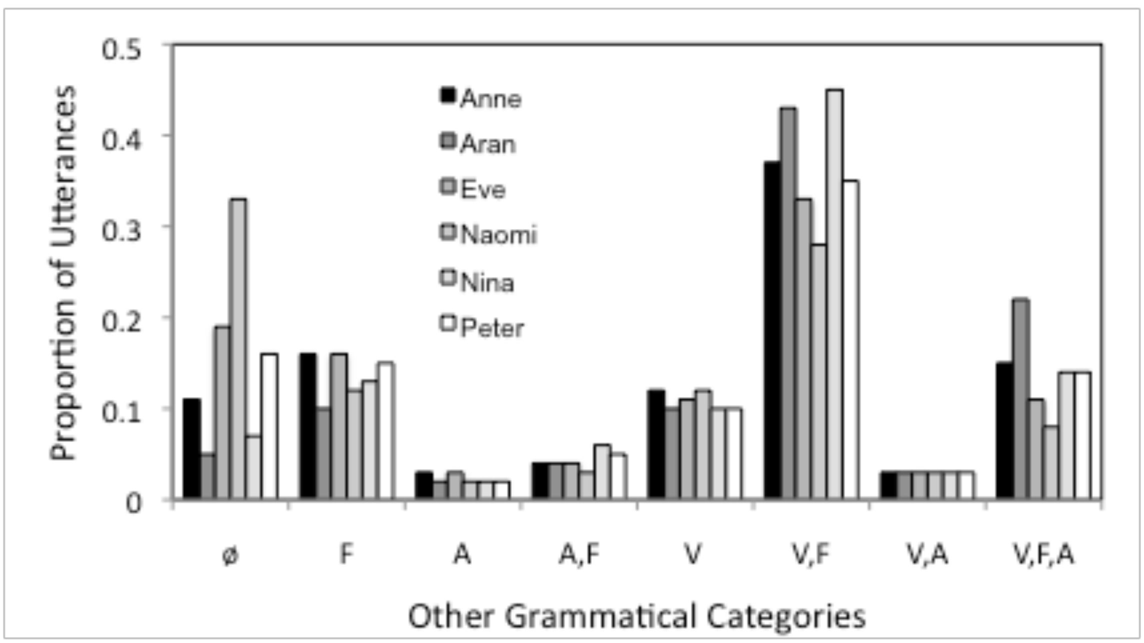


Figure 2. Proportion of utterances containing one or more nouns and function words, adjectives/adverbs, and verbs. F: utterance contains ≥1 function words; A: utterance contains ≥1 adjectives/adverbs; V: contains ≥1 verbs.

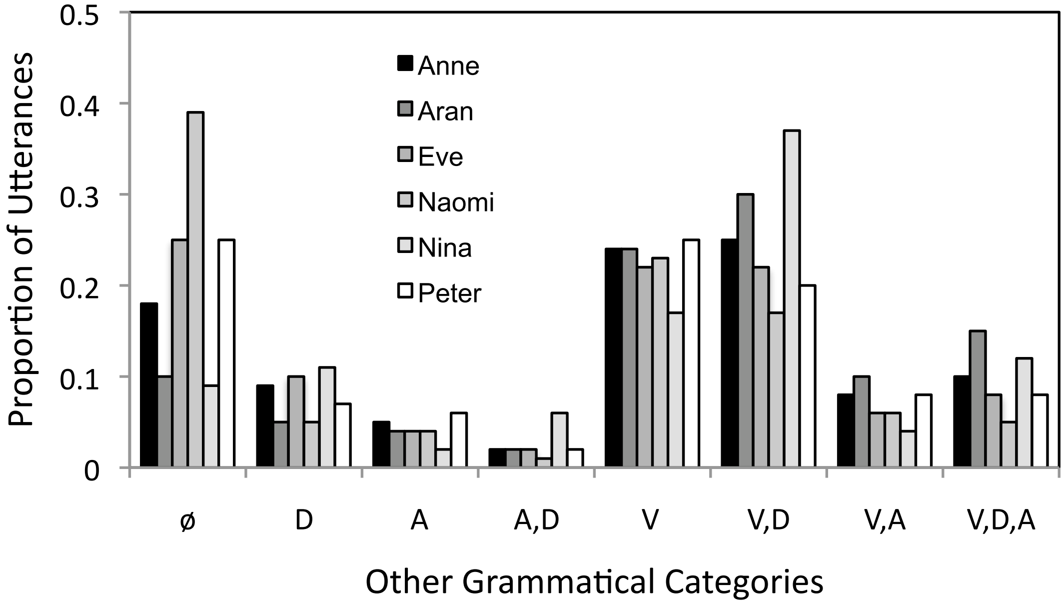


Figure 3. Proportion of utterances containing one or more nouns and determiners, adjectives/adverbs, and verbs. D: utterance contains ≥1 determiners; A: utterance contains ≥1 adjectives/adverbs; V: contains ≥1 verbs.

a potential referent, or no nouns at all (in these cases, an average of 49% of the cases contained at least one pronoun).

Figure 2 indicates for the utterances with one or more nouns the proportion that also contain verbs, adjectives and adverbs, and function words. An average of 15% of the utterances containing one or more nouns contain no other words and may have provided ideal information about the pairing of the word with an object in the child's environment. Almost half (46.7%) of these utterances containing only nouns contained proper nouns, the ideal scenario word-referent pairings, where there is just one noun and no other potential referring word in the utterance, comprised approximately 8% of the noun only utterances. However, more common is when a noun is accompanied by at least one other word that could not be a potential referring word. Sixty-five percent of utterances containing one or more nouns also contain at least one verb, 24% contain at least one adjective or adverb, and 69% contain at least one function word. Of particular relevance is that when the utterance contains at least one noun and another content word, a function word is also present in the utterance in 77.3% of cases.

As we are most interested in the distribution of determiners indicating noun categories, we also analysed the proportion of utterances with one or more nouns that also contain verbs, adjectives and adverbs, and determiners. The results are shown in Figure 3. Consistent with the results presented in Figure 2, when there is at least one noun and another content word in the utterance then it also contains a determiner in 52.6% of cases. This suggests that other function words, which can also provide information about grammatical categories, are also prevalent when multiple content words are present in the utterance. For instance, for utterances containing at least one noun and one other content word, 48% of these utterances also contain a preposition, which provides highly reliable distributional information about the noun versus verb distinction in child-directed speech (Monaghan et al, 2007).

The corpus analyses confirmed that children are exposed to situations where several nouns, potentially referring to objects in the environment, occur in the same utterance. However, more frequently, children hear an utterance that contains several words other than the noun. So, the task of cross-situational learning requires learning which of the words in a multi-word utterance may relate to an object in the environment and which do not. The words to be rejected may either co-occur with a particular object (in the case of certain verbs, adverbs, and adjectives), or may co-occur with an object but also with many other objects (function words). We next assess the position that these potentially informative function words take in the utterances containing multiple content words.

Table 2. Proportion of utterances containing one or more nouns or a single noun preceded by determiners comparing utterances with/without a verb. The Z score of the comparison for each corpus is derived from Mann-Whitney's U test. $Z_{-pn}$ refers to the Z score of the analyses omitting proper nouns from the analysis.

| | One or More Nouns | | | | One Noun | | |
|---|---|---|---|---|---|---|---|
| | With Verb | Without Verb | Z | $Z_{-pn}$ | With Verb | Without Verb | Z |
| Anne | .46 | .35 | 12.54*** | 12.78** | .49 | .37 | 9.92*** |
| Aran | .46 | .42 | 4.11*** | 4.59** | .48 | .44 | 2.85** |
| Eve | .40 | .26 | 14.96*** | 10.27** | .42 | .27 | 11.82*** |
| Naomi | .39 | .21 | 16.65*** | 12.85** | .42 | .20 | 14.71*** |
| Nina | .49 | .41 | 9.05*** | 7.89** | .53 | .42 | 8.68*** |
| Peter | .53 | .34 | 19.15*** | 15.95** | .55 | .35 | 16.58*** |

** $p < .005$; *** $p < .001$

To test the hypothesis that in more complex utterances contextual information may provide more reliable constraints to the nouns in speech, we compared the proportion of nouns preceded by a determiner when a verb also occurred in the utterance to when no verb occurred in the utterance. Within each corpus, the proportions of utterances with versus without verbs were compared using Mann-Whitney's U. For each corpus (results shown in Table 2), children were more likely to hear nouns preceded by determiners when there was also a verb – a word not referring to an object – in the utterance. Grammatical constraints therefore assist in highlighting the potential object-referring word, and this reliability increases if the sentence contains content words from the noun and verb categories. It may be the case that utterances containing no content words other than nouns are distinct in terms of their communicative function – for instance, they may consist of proper nouns to a greater degree than nouns used in utterances containing other content words. We repeated the analyses omitting proper nouns, and found that the results were very similar, as shown in the $Z_{-pn}$ column of Table 2.

In order to establish whether this effect of increased reliability of determiner use was due to sentences containing verbs tending to contain more nouns (therefore increasing the chance of a noun being preceded by a determiner), we repeated the analyses restricting the utterances assessed to those containing only one noun. The results were very similar, and are shown in the rightmost columns of Table 2. The results suggest that if a verb is also present in an utterance then a noun is more likely to be preceded by a determiner. We repeated the results testing whether nouns were more likely to be preceded by determiners if the utterance contained one or more adverbs or verbs, to determine whether introducing another type of content word resulted in a similar pattern of results. The results were very similar, see Table 3 leftmost columns.

However, increasing the complexity of the utterance in other ways may reduce the reliability of direct contextual information for identifying potential referring words in the utterance. In particular, adjectives also add complexity, but are likely to reduce the proportion of nouns preceded by determiners, as they tend to occur in a noun phrase preceding the noun and succeeding a determiner. We repeated the analyses comparing utterances containing at least one noun and at least one verb, adverb, or adjective to those containing at least one noun and no other content words. The results are shown in the rightmost columns of Table 3. The greater reliability of determiners directly preceding nouns for more complex utterances was maintained in only 4 of the 6 corpora. The Aran and Nina corpora were those that contained the highest proportion of utterances containing adjectives – 13.2% and 12.0%, compared to the 8.0% to 9.2% proportion of utterances for the other 4 corpora. Thus, increasing complexity of utterances may in some cases reduce the reliability of adjacent contextual information (see Mintz, 2003, and St Clair et al., 2010, for discussion and resolution of this issue).

Table 3. Proportion of utterances containing one or more nouns and with/without a verb or adverb that are preceded by a determiner, and proportion of utterances containing one or more nouns and with/without a verb or adverb or adjective that are preceded by a determiner. The Z score of the comparison for each corpus is derived from Mann-Whitney's U test.

| | Verb/Adverb | | | Verb/Adverb/Adjective | | |
|---|---|---|---|---|---|---|
| | With | Without | Z | With | Without | Z |
| Anne | .46 | .35 | 12.26*** | .44 | .39 | 5.53*** |
| Aran | .46 | .42 | 3.50*** | .45 | .49 | -2.94* |
| Eve | .39 | .25 | 14.73*** | .38 | .27 | 10.78*** |
| Naomi | .40 | .19 | 18.51*** | .38 | .20 | 16.78*** |
| Nina | .50 | .39 | 10.50*** | .47 | .47 | -.01 |
| Peter | .53 | .33 | 19.40*** | .50 | .36 | 12.80*** |

* $p < .01$; *** $p < .001$

Finally, to test whether there was a proportional relationship between complexity of the utterance and the reliability of distributional information for nouns, we correlated the utterance length and the proportion of nouns within the utterance that were preceded by determiners for each child-directed speech corpus. For each corpus there was a significant correlation between utterance length and reliability of the determiners preceding nouns, ranging from $r = .35$ for the Aran corpus to $r = .41$ for the Anne corpus. For all 6 corpora, $p < .001$. Thus, though a large proportion of adjectives in utterances may have reduced reliability of the contextual cues to referring words, overall, the utterance length was related to how reliable referring words were preceded by determiners in the child directed speech, indicating a direct relationship between utterance complexity and reliability of nouns being preceded by determiners.

We next tested whether learners were able to learn cross-situational statistics in the presence of non-referring words, and, furthermore, whether the contextual information observed in child-directed speech functioned as a distributional constraint to assist learning of word-referent mappings.

## Study 2. Cross-situational learning with and without contextual constraints

*Method*

*Participants*. 45 students at Lancaster University, 17 males and 28 females with mean age 20.7 years (SD = 2.1), volunteered for participation in the study. All reported speaking English as their first language and had normal vision and hearing.

*Materials*. For the objects, we selected 10 geometric shapes printed in black on a grey background, taken from Fiser and Aslin (2002). For the words, 22 nonsense words were recorded by a female native-English speaker in a neutral tone. The referring and non-referring words were bisyllabic and were selected randomly from a set of 20, shown in the Appendix. In addition, there were two monosyllabic "function words", which were also non-referring words, but could provide distributional information as to the referring word in the language, according to whether the language was structured to include such constraints. Bisyllabic words were 900ms in length, and monosyllabic words were 500ms.

Pictures were randomly paired with 10 of the bisyllabic words for each participant, with the remaining 10 bisyllabic words as non-referring words. There were three conditions. The first was the internal structure language with contextual constraints, where one of the monosyllabic function words marked the referring words, and the other monosyllabic word marked the non-referring words. Which function word marked the referring words was randomised across participants. The second condition was the no internal structure language, where the pairing between the function words and the referring and non-referring words was randomised, such that the function words did not indicate which word was the referring word. The comparison between these conditions enabled a test of whether, for languages of similar complexity, participants could exploit the distributional information that indicated grammatical categories of words in the language in learning to pair words with referents. The final condition tested whether the grammatical structure of the language facilitated learning over a simpler language where the function words were omitted. For this no marker word condition, participants heard just the referring and the non-referring word in each sentence. Participants were randomly assigned to each condition.

*Table 4. Cross-situational statistics for the materials in Study 2.*

| A | B | $p$(B\|A) |
|---|---|---|
| Referring word | Target object | 1.0 |
| Referring word | Non-target object | .11 |
| Non-referring word | Target object | .10 |
| Non-referring word | Non-target object | .10 |
| Function word* | Target object | .10 |
| Function word* | Non-target object | .01 |
| Target object | Referring word | .50 |
| Non-target object | Referring word | .11 |
| Target object | Non-referring word | .10 |
| Non-target object | Non-referring word | .10 |
| Target object | Function word* | 1.0 |
| Non-target object | Function word* | 1.0 |

* For internal structure and no internal structure conditions only.

The cross-situational statistics for the stimuli are shown in Table 4. They indicate that the referring word and target object pairing is the only pairing that should be learned by participants, as this is the only informative correspondence in the data set. The statistics shown in Table 4 apply from the first block of training, as the cross-situational statistics were controlled within each block as well as over the whole experiment. We predicted that the internal structure language would be learned more accurately than the no internal structure language. In addition, if the greater complexity, but increased constraints of the internal structure language was sufficient to overcome the greater simplicity, but fewer constraints, of the no marker word language, then an increase in accuracy should also be seen for the internal structure language over the no marker word language. We also predicted that the effect of internal structure would result in an increasing advantage with exposure to the language as there was likely to be a period of learning the constraints of the language structure prior to applying it to the word-object mappings, resulting in a significant interaction between training block and language condition. We also predicted that the performance in the internal structure condition would be significantly better than chance performance at an earlier stage than the no internal and no marker word language conditions.

*Procedure*. For each trial, the participant saw two pictures presented on the computer screen. After 500ms they heard a sentence. For the internal structure and no internal structure languages, this was composed of four words: function word, referring word, function word, non-referring word (the order of the referring and non-referring words was randomised across trials). For the no marker word condition, the sentence was composed of just two words: the referring word and the non-referring word, with order randomised across trials. The target and alternate picture were randomised for screen position (left or right) across trials. Following sentence presentation, participants were instructed to press “1” on the keyboard if the left picture was described, and “2” if the right picture was described. 1000ms after the participant’s response, the next trial began. An example trial is shown in Figure 4. The instructions to the participant provided information to choose a single referent, corresponding to the fact that there are frequent occasions when only a single referent is referred to in speech, as demonstrated in Yu and Ballard's (2007) analyses of word-referent mappings in naturalistic situations. It is important to note that the instructions did not provide information to the learner about the roles of the words in the utterance.

Trials were presented in blocks of 30. Each picture occurred 3 times as the target, and 3 times as the non-referring picture, and the screen position of the target and non-referring picture was pseudo-randomised so that the target occurred an equal number of times as the left and as the right picture. Words occurred an equal number of times in each block. After 4 blocks of training, participants were given the opportunity to rest. The experiment concluded after 8 blocks of training. Participants were exposed to each word-referent mapping 24 times.

*Results and discussion*

For each training block, we recorded the proportion of trials in which the correct target picture was selected, and the proportion correct was then arcsine-transformed. The results for the untransformed proportion correct are shown in Figure 5. In order to test whether learning was better than chance in the presence of non-referring words, we compared the performance of all three conditions to chance level (arcsine(.5) = .524). For the internal structure language, performance was better than chance for blocks 5 to 8 (all $p < .01$, with

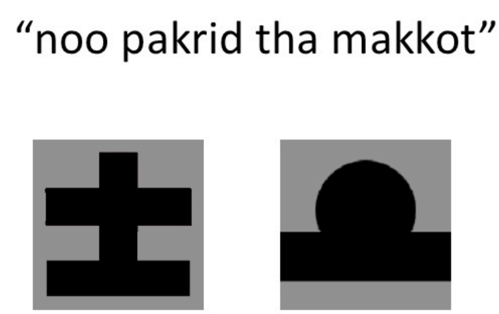


Figure 4. Example of one trial. For this participant in the internal structure language, "noo" indicates that "pakrid" is the referring word, and the left shape is the correct target. "makkot" never refers to a shape, and the right shape is referred to by another word not spoken at this time.

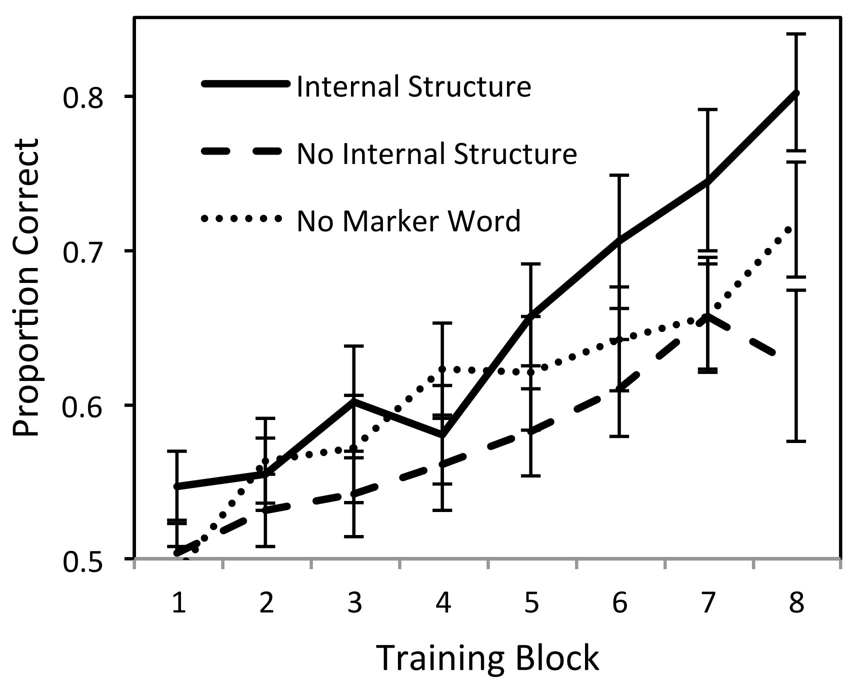


Figure 5. Proportion of correct pictures selected in each training block for Study 2. Error bars show Standard Error of the Mean.

Bonferroni correction) but was not significantly better than chance for blocks 1 to 4 (after Bonferroni correction). For the no internal structure language, performance was better than chance for blocks 6 and 7, $p < .05$. For the no marker word language, performance was better than chance for blocks 4, 6, 7, and 8, $p < .01$. Over all blocks, the mean correct for the internal structure language was .65 (SD = .10), which was significantly better than chance, $t(14) = 5.41$, $p < .001$. For the no internal structure language, mean correct was .58 (SD = .09), which was again better than chance, $t(14) = 3.34$, $p < .005$. For the no marker word condition, proportion correct over all blocks was .61 (SD = .09), which was also better than chance, $t(14) = 4.48$, $p < .001$. Thus, the presence of non-referring words as found in child-directed speech did not prevent learning of the word-referent mappings in any of the conditions. Furthermore, the internal structure language resulted in better than chance performance at an earlier stage than the no internal structure language (from block 5, compared to from block 6), though the no marker word condition also resulted in performance better than chance at a still earlier stage (from block 4, though this dropped below significance for block 5) which we suggest was likely due to the reduced complexity of the utterances in this condition (2 words instead of 4). Hence, our prediction that the internal structure language would be learned earlier than the other conditions was only partially supported.

We conducted an ANOVA on arcsine transformed proportion correct, with training block (1-8) as a within-subjects factor and language condition (internal structure, no internal structure, no marker word) as a between-subjects factor. For the main effect of block, Mauchley's sphericity test was significant, and so we used the Greenhouse-Geisser correction. There was a main effect of block, $F(3.82, 160.58) = 26.10$, $p < .001$, $\eta^2 = .38$, with improved learning across the experiment, as indicated by a significant linear contrast effect, $F(1, 42) = 58.69$, $p < .001$, $\eta^2 = .58$. There was no significant main effect of language, $F(2, 42) = 2.51$, $p = .09$, $\eta^2 = .11$. However, the interaction between block and language was significant, $F(7.65, 160.58) = 2.09$, $p < .05$, $\eta^2 = .09$, indicating that the difference between the languages altered as training proceeded. To test the pattern of this interaction across blocks for the different comparisons, we performed linear and quadratic contrasts. There was a marginally significant effect of the quadratic contrast, $F(2, 42) = 3.17$, $p = .05$, $\eta^2 = .13$, but the linear contrast was not significant, $F(2, 42) = 2.74$, $p = .08$, $\eta^2 = .12$. This suggests that the middle blocks of training differed from the first and last blocks in terms of the distinction among conditions.

To determine the source of the interaction, we conducted planned comparisons between pairs of the language conditions. For the comparison between internal structure and no internal structure conditions, there was a main effect of condition, $F(1, 28) = 4.68$, $p < .05$, $\eta^2 = .14$, with better performance for the internal structure language. There was a significant main effect of block, $F(3,11, 87.02) = 18.51$, $p < .001$, $\eta^2 = .40$, and a significant interaction between condition and block, $F(3.11, 87.02) = 3.00$, $p < .05$, $\eta^2 = .10$, with increasing advantage for the internal structure language with additional training. For blocks 5 and 6, there was a marginally significant difference between the internal and no internal structure languages, $p = .08$ and $p = .07$, respectively, and this comparison was significant only for block 8, $p < .05$. The linear contrast for the interaction was marginally significant, $F(1, 28) = 4.17$, $p = .05$, $\eta^2 = .13$, and the quadratic contrast was not significant, $F(1, 28) = 3.36$, $p = .08$, $\eta^2 = .11$, which indicates that the difference between languages with and without internal structure increases with training time.

For the comparison between internal structure and no marker words, there was no significant main effect of condition, $F(1, 28) = 1.54$, $p = .23$, $\eta^2 = .05$. There was a significant main effect of block, $F(3.95, 110.45) = 22.24$, $p < .001$, $\eta^2 = .44$, and a significant interaction, $F(3.95, 110.45) = 2.22$, $p < .05$, $\eta^2 = .07$. There was a marginally significant learning advantage for the internal structure condition in the first and last block of training, $p = .06$, and $p = .09$, respectively, which was reflected in the significant quadratic contrast for the interaction, $F(1, 28) = 4.71$, $p < .05$. $\eta^2 = .14$. The linear contrast for the interaction was not significant, $F(1, 28) = 2.66$, $p = .11$, $\eta^2 = .09$, indicating that there was an initial and final advantage for the internal structure language across blocks. At block 1, the internal structure condition was not significantly different from chance level after Bonferroni correction , but was significant before correction, $t(14) = 2.16$, $p = .04$, suggesting an early boost to learning in this condition.

For the comparison between no internal structure and no marker words, there was no significant main effect of condition, $F(1, 28) = 1.04$, $p = .32$, $\eta^2 = .04$, a significant main effect of block, $F(4.28, 119.69) = 12.51$, $p < .001$, $\eta^2 = .31$, and no significant interaction, $F < 1$. The linear and quadratic contrasts for the interaction were also not significant, both $F < 1$.

The results support the hypotheses generated by the corpus analysis, that the grammatical structure of speech assists in acquisition of word-referent mappings. This was the case when the complexity of the language was kept stable, but varied whether the complexity provided cues to the roles of words in the sentence. The advantage of internal structure was also observed early and later in training in comparison to a simpler language containing no function words.

## General Discussion

Quine (1960) highlighted the difficulty of learning word-referent mappings due to the many possible pairings that can be made between words and potential referents. In his famous example of a language learner observing a scene of a rabbit running and hearing a native speaker's utterance, "gavagai", the word could refer to a rabbit, to the rabbit's fur, the rabbit's action, or the general beauty of the scene. Uncovering the constraints that language learners apply when they are acquiring words (e.g., Markman, 1990) indicates ways in which this multitude of possibilities is reduced.

Yet, Quine's example omits an additional complexity to the word learning situation facing learners acquiring word-referent mappings: utterances in child-directed speech contain multiple words. As shown by our corpus analyses, not only are there multiple possible referents for each word, but also multiple words in each utterance, increasing the difficulty from a one-to-many mapping to a many-to-many mapping. The corpus analyses confirmed that a common situation in child-directed speech is one where there are multiple nouns present in speech, and presumably, as in the detailed analyses presented by Yu and Ballard (2007), also multiple possible objects to which they may refer. The cross-situational learning shown to be effective in previous studies (Smith & Yu, 2008; Yu & Smith, 2007) reflects this situation of multiple nouns present in speech.

However, the corpus analyses also indicated that in the majority of cases where a noun occurs, other words also occur. The usual situation for children's exposure to language is of complex, multi-word utterances comprising multiple grammatical categories. Yet, for more complex utterances that contain at least one noun and verb, we discovered that distributional information marking the potential referring word – the noun – is more likely than when the utterance contains no verb. Reliable information is especially useful for the child when it is important to constrain the potential word-object mappings from a larger range of alternatives. Our corpus analyses have only scratched the surface of the possible complexity available to assist the child in learning word-referent mappings. For instance, we limited our analyses to single utterances which underestimate the information available across utterances, through repeated reference to an object using its noun label or deictic pronouns. Yet, even by considering utterances in isolation, there are grammatical cues that provide powerful constraints for learning reference.

Our experimental study showed that cross-situational learning is robust to the presence of multiple non-referring words. We also showed that the presence of these non-referring words provides additional useful constraints for determining word-referent mappings in language learning. If the hearer can determine, from the internal structure of the utterance, that "gavagai" is a noun, then that limits the possible semantic properties of the referent – it cannot now be the action of the rabbit or the general beauty of the scene. Our experimental work confirms the theoretical claims of Gleitman (1990) who discussed the use of syntactic properties that are required to constrain the possible referents for verb learning. We have shown that these constraints can apply as additional scaffolds to noun learning, and we hypothesise that referents for other grammatical categories of words may also be supported by analogous language-internal features. For instance, verbs also can be reliably cued by distributional information. Monaghan et al. (2007) found that the most frequent 20 words in the language provided accurate information for classifying both nouns and verbs in a range of languages. Pronouns and prepositions were particularly useful for indicating that a verb is likely to be the next word in the utterance.

It is an interesting issue as to whether such distributional information providing syntactic information is learned prior to, simultaneously with, or following learning of the first word-referent mappings in the language. Our study does not provide a direct answer to this, though the experimental study shows that early in training for our participants, the grammatical structure of the language begins to exert a positive influence on learning before many of the word-referent mappings had been learned, as evidenced by the quadratic contrast interaction between block and the internal structure and no marker word condition, demonstrating an early effect, and the significant interaction between block and internal versus no internal structure conditions. Similarly, Yu (2006) proposed that syntactic information provides a useful bootstrap to be used in conjunction with learning associations between individual words and their referents. This suggests that interactions between meaning and grammatical structure are likely to be important early in vocabulary acquisition.

The three conditions that we tested addressed the issue of whether the potential distributional information introduced by function words in speech assisted in acquiring word-referent mappings, or whether the effect of no internal structure actually inhibited learning. Comparing the internal structure and no internal structure conditions demonstrated a learning advantage for internal structure, but this may have been because participants in the no internal structure condition were actually impaired in their learning.

The no marker words condition provided a strong test of this effect. Despite the simpler structure of the no marker words condition, with only 2 words to determine in terms of their role in describing the scene instead of 4, there was an initial and final advantage for the language with internal structure. Hence, presence of function words appears to be beneficial for learning. Furthermore, there was no significant difference in performance between the no internal structure condition and the no marker words condition, thus the presence of function words that did not cue grammatical roles of words or their absence entirely, resulted in similar patterns of learning. This raises the question about why the learning advantage should be observed early and late in training. Figure 5 suggests that the internal structure provides increasing benefit as training proceeds – with the difference in mean accuracy increasing from blocks 5 through to 6, so extended training appeared to increase the learning advantage of internal structure. Yet, the early advantage of function words may have enhanced the identification of a subset of the referring words.

Monaghan and Mattock (2009) tested a language similar to the internal structure language used in the current study, though with 6 instead of 10 word-object mappings to be learned, and compared it to a no marker word condition. Though there was again no overall advantage of the internal structure versus the no marker words condition for this smaller language, a reanalysis of the data revealed a significant quadratic contrast in the interaction between block and condition, similar to that found in the current study. This provides convergent evidence that the advantage of the language with internal structure, early and late in training, is a robust effect promoting learning of word meanings. This was consistent with our expectation that the effect of internal structure would increase as training proceeded, but we did not expect an early advantage for the internal structure over the no marker word condition. Hence, the internal structure provides some initial advantage for constraining potential referring words, as well as a later advantage as performance for the internal structure language diverges from the other two language conditions. In the intervening blocks, the variability in responses between conditions may have obscured the potential advantage for the internal structure language which only re-emerges as accuracy increases to a sufficient level in later blocks of training (see Figure 5). Simulation studies have confirmed that learning a large set of word-referent pairs may even be easier than learning a smaller set (Yu, 2006), and we predict that as the language increases in vocabulary size and grammatical complexity, the distributional constraints will increase in their beneficial effect for word learning.

The grammatical constraints useful for constraining potential referring words and referents are likely to be available to children at an early stage of language learning. At 7-9 months, infants can detect function words in continuous speech (Höhle & Weissenborn, 2003) and can use function words to segment a noun following the function word (Shi & Lepage, 2008). By 13-months, children are able to distinguish subtle phonological variations to function words (Shi, Werker, & Cutler, 2006), and by 14-16 months, infants are sensitive to distributional information that determiners precede noun categories (Höhle et al., 2004; Shi & Melançon, 2010), and by 18 months, infants are producing two-word combinations, which is a conservative measure for the onset of syntactic knowledge (Lidz, Waxman, & Freedman, 2003; Saffran & Wilson, 2003). Infants are therefore sensitive to relationships between function words and content words at ages critical for deriving word-object mappings. We suggest that this nascent distributional knowledge has an impact on learning word meanings.

The correspondence between the child-directed speech corpus analysis and the experimental study in terms of distributional cues available and potentially enhanced in word learning situations provides insight into language acquisition processes, and is in accordance with other studies of adult artificial language processing that exemplify similar patterns of behaviour to infants (e.g., Smith & Yu, 2008 and Yu & Smith, 2007). However, there remains the possibility that the use of distributional cues by participants may have been partially influenced by the learners' first language, English. One way to address this in future research is to repeat the experimental study but vary the order in which the function words and the referring and non-referring words occur. Though participants are more likely to learn from preceding function words than succeeding function words, based on previous research with English (Frigo & MacDonald, 1998, we predict that participants would be sensitive to function words following the referring words even though the structure counteracts that of English (though does still relate to the suffixing preference in English, St Clair, Monaghan, & Ramscar, 2009).

Another limitation of the current study is that the testing method used in our experimental study did not permit distinguishing whether participants learned the mapping from the referring word to the object or the mapping from both the marker word and referring word to the object. Frequently co-occurring words tend to be processed as unanalysed chunks by children (Bannard & Matthews, 2010). However, the fact that the marker word occurred in every utterance in the experiment together with the computational work showing the possibility of decoupling function words from referents

(Yu & Ballard, 2007), suggests to us that the learning was driven by learning of the referring word to object mapping.

In conclusion, word learning situations present ambiguity not only because of the range of possible referents for each word, but also because there are multiple words in each situation only some of which map onto objects in the environment. We have shown that these multi-word utterances provide valuable language-internal constraints to supplement the numerous other language-external constraints (Akhtar, 2002; Baldwin, 1993; Gleitman, 1990; Siskind, 1996; Tomasello, 2003; Yu, 2006; Yu & Smith, 2007) that are critical for language learning.

*References*


Akhtar, N. (2002). Relevance and early word learning. *Journal of Child Language, 29*, 677-686.

Aslin, R.N., Woodward, J., LaMendola, N., & Bever, T.G. (1996). Models of word segmentation in fluent maternal speech to infants. In J.L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition (pp.117-134)*. Mahwah, NJ: Lawrence Erbaum Associates.

Baldwin, D. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29*, 832-843.

Bannard, C., & Matthews, D. E. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19*, 241-248.

Bloom, L., Hood, L., & Lightbown, P. (1974). Imitation in language development: If, when and why. *Cognitive Psychology, 6*, 380-420.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Fiser, J. & Aslin, R.N. (2002). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 458-467.

Frigo, L., & McDonald, J. L. (1998). Properties of phonological markers that affect the acquisition of gender-like subclasses. *Journal of Memory and Language, 39*, 218–245.

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

Gogate, L., Walker-Andrews, A.S., & Bahrick, L.E. (2001). Intersensory origins of word comprehension: An ecological-dynamic systems view. *Developmental Science, 4*, 1-37.

Höhle, B. & Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science, 6*(2), 122-127.

Höhle, B. Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M. (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy, 5*(3), 341-353.

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R.M. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Society for Research in Child Development Monograph Series*. Chicago: University of Chicago Press.

Horst, J. S., Scott, E. J. & Pollard, J. A. (2010). The role of competition in word learning via referent selection. *Developmental Science, 13*, 706-713.

Katz, N., Baker, E., & MacNamara, J. (1974). What's in a name? On the child's acquisition of proper and common nouns. *Child Development, 45*, 269-273.

Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: Experimental evidence for syntactic structure at 18 months. *Cognition, 89*, 295-303.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Volume 2: The database, 3rd edition*.

Markman, E.M. (1990). Constraints children place on word learning. *Cognitive Science, 14*, 57-77.

Mintz, T.H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*, 91-117.

Mintz, T.H., Newport, E.L., & Bever, T.G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*, 393-424.

Monaghan, P., & Christiansen, M.H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language, 37*, 545-564.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The Phonological Distributional Coherence Hypothesis: Cross-linguisitic evidence in language acquisition. *Cognitive Psychology, 55*, 259-305.

Monaghan, P. & Mattock, K. (2009). Cross-situational language learning: The effects of grammatical categories as constraints on referential labelling. *Proceedings of the 31st Cognitive Science Society Conference*. Mahwah, NJ: Lawrence Erlbaum.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

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

Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent-child discourse. In K. E. Nelson (Ed.), *Children's Language* (4th ed., pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum Associates.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4*(2), 273-283.

Sagae, K., MacWhinney, B., & Lavie, A. (2004). Automatic parsing of parental verbal input. *Behavior Research Methods, Instruments and computers, 36*, 113-126.

Shady, M., Gerken, L. A., & Jusczyk, P. W. (1995). Some evidence of sensitivity to prosody and word order in ten-month-olds. In D. MacLaughlin & S. McEwen (Eds.), *Proceedings of the 19th Annual Boston University Conference on Language Development* (Vol. 2, pp. 553-562). Somerville, MA: Cascadilla Press.

Shi, R. & Lepage, M. (2008), The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science, 11*, 407–413.

Shi, R. & Melançon, A. (2010). Syntactic categorization in French-learning infants. *Infancy, 15*, 517–533.

Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy, 1*0(2), 187-198.

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

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

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

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

St Clair, M.C., Monaghan, P., & Christiansen, M.H. (2010). Learning grammatical categories from distributional cues: Flexible frames for language acquisition. *Cognition, 116*, 341-360.

St. Clair, M. C., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science, 33*, 1317–1329.

Suppes, P. (1974). The semantics of children's language. *American Psychologists, 29,* 103-114.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language, 28*, 127-152.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language, 59*, 114-132.

Yu, C. (2006). Learning syntax-semantics mappings to bootstrap word learning. In R. Sun (Ed.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 924-929). Mahwah, NJ: Lawrence Erlbaum Associates.

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

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

*Acknowledgements*

Thanks to Alastair Smith for assisting in data collection and to Emma Threadgold for recording the speech stimuli. Thanks also to three anonymous reviewers who suggested several of the corpus analyses.

*Appendix: Speech stimuli*

Bisyllabic words used as referring and non-referring words in Study 2:

*barget, bimdah, chelad, dingep, fisslin, goorshell, haagle, jeelow, kerrwoll, limeber, makkot, nellby, pakrid, rakken, shooglow, sumbark, trepier, vinnoy, wiertat, zawyer*

Monosyllabic words used as function words in Study 2:

*tha, noo*