Examining the role of lexical frequency in the acquisition and processing of sentential complements

Evan Kidd [a,*], Elena Lieven [b], Michael Tomasello [b]

[a] *Max Planck Child Study Centre, The University of Manchester, United Kingdom*
[b] *Max Planck Institute for Evolutionary Anthropology, Leipzig, United Kingdom*

Abstract

We present empirical data showing that the relative frequency with which a verb normally appears in a syntactic construction predicts young children's ability to remember and repeat sentences instantiating that construction. Children aged 2;10–5;8 years were asked to repeat grammatical and ungrammatical sentential complement sentences (e.g., 'I think + S'). The sentences contained complement-taking verbs (CTVs) used with differing frequencies in children's natural speech. All children repeated sentences containing high frequency CTVs (e.g., *think*) more accurately than those containing low frequency CTVs (e.g., *hear*), and made more sophisticated corrections to ungrammatical sentences containing high frequency CTVs. The data suggest that, like adults, children are sensitive to lexico-constructional collocations. The implications for language acquisition are discussed.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Sentential complements; Language acquisition; Language processing; Frequency; Syntax

Research in adult sentence processing has consistently shown that the parser is guided by lexical preferences at points of syntactic ambiguity (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Holmes, Stowe, & Cupples, 1989; Trueswell, Tanenhaus, & Kello, 1993). In languages like English, information encoded on the lexical entry for verbs provides a wealth of information about the upcoming structure of a sentence. Consider sentences (1) and (2).

(1) The professor knew the student would complete her assignment.
(2) The professor knew the student from his job as course advisor.

The verb *know* subcategorizes for either a sentential complement (1) or a direct object (2). Thus, there is ambiguity at *the student* as to the continuation of the sentence which is not resolved by

* Correspondence to: School of Psychological Sciences, The University of Manchester, Oxford Road, M13 9PL, United Kingdom Tel.: +44 161 275 2578; fax: +44 161 275 8587.
E-mail address: evan.j.kidd@manchester.ac.uk (E. Kidd).

argument structure information. Sentence processing studies with adult participants have shown that individual verbs prefer one continuation over the other. Holmes et al. (1989), for instance, showed that the lexical preferences of verbs affected reading times: verbs that most often occurred in sentential complement constructions biased a sentential complement reading, and verbs that most often occurred in simple transitive structures biased a direct object reading (see also Garnsey et al., 1997; Trueswell et al., 1993). These findings suggest a central role for lexical access in structure building. In particular, they identify the relative frequency with which a verb occurs in a particular constructional pattern as playing a prominent role in guiding structural choice.

On the basis of several analyses of large corpora, Lapata, Keller, and Schulte im Walde (2001) argued that such lexical preferences are records of prior linguistic experience. That is, they argued that the parser pursues the syntactic analysis that is best attested in the input. Arguing for a key role for experiential processes in language processing raises at least two developmental questions. First, at what point in development does frequency play a role in children's understanding and use of language? Second, to what extent does frequency drive acquisition? We consider these questions within the framework of the 'emergentist' approach to language acquisition.

According the emergentist approach, children induce a grammar using general-cognitive learning mechanisms. Researchers working within this approach emphasize the restricted nature of children's early linguistic knowledge: children break into the grammar of their language by learning high frequency predicate-argument exemplars as sentence level form-meaning pairings (e.g., Bates & MacWhinney, 1989; Goldberg, 2005; MacWhinney, 1998; Tomasello, 2003). Verbs become *entrenched* in constructions; for example, the verb *know* would become entrenched in the finite sentential complement construction (Verb-Sentence, as in 1) and the transitive construction (Verb-Object, as in 2). The more a verb appears in a construction the more strongly it will become entrenched in that constructional pattern. The approach argues that children converge on the adult grammar by making analogies over lexically specific knowledge. The basis for analogy is far from clear; however, one proposal is that high type-token frequency facilitates abstraction (see Goldberg, Casenhiser, & Sethuraman, 2004; Goldberg, 2005; Tomasello, 2003). Frequency is therefore given a prominent role in acquisition, and there is now a growing body evidence to suggest that children are sensitive to lexical frequency information.

Theakston, Lieven, Pine, and Rowland (2004) showed that input frequency best predicted young children's early use of verbs in multiple syntactic environments. In an elicited production experiment, Brooks, Tomasello, Dodson, and Lewis (1999) showed that children were less likely to provide argument structure generalization errors when they were tested on high frequency than on low frequency verbs (see also Theakston, 2004). Additionally, Matthews, Lieven, Theakston, and Tomasello (2005) showed that 2- and 3-year-old children were more likely to correct an ungrammatical sentence when it contained a high frequency verb. Such results support the argument that lexical frequency is important, but clearly more research needs to be conducted in order to fully elucidate the role of frequency in acquisition. Currently a majority of this work has investigated verb frequency effects in the acquisition of the transitive and intransitive constructions; we report on a study investigating the role of lexical frequency in the acquisition and subsequent processing of finite sentential complements. We briefly discuss the literature on the acquisition complement clause constructions to situate our work.

Complement clause constructions are complex sentences where a sentential constituent functions as an argument of a main verb (as in 1). Most research investigating the acquisition of complement clauses has used naturalistic data. Children begin to produce sentential complement constructions soon after their second birthday. Their early knowledge of complementation appears to be lexically specific; children's productions are initially restricted to a small number of highly

frequent verbs that take sentential complements. Bloom, Rispoli, Gartner, and Hafitz (1989) reported on the production of *wh-* and sentential complements licensed by four highly frequent complementizer verbs: *think*, *know*, *see*, and *look* [*at*] by four children followed longitudinally between the ages of 2- and 3-years. They argued that that these children's knowledge of the syntax of complements was idiosyncratically constrained by individual complementizer verbs. For instance, for the two verbs *think* and *know*, *think* exclusively occurred with *that-* complements, whereas *know* occurred most often with *wh*-complements. While the children were syntactically productive within the complement clause, their matrix clauses were highly formulaic; in particular, the subject of the sentence was restricted largely to first and second person pronouns (although this depended on the individual verb) and the main verb was rarely inflected. In contrast, children were much more productive within the subordinate clause.

Diessel and Tomasello (2001, see also Diessel, 2004) analysed naturalistic data from seven children followed longitudinally, from age 1;2 to 5;2, enabling them to study children's use of a larger range of complementizer verbs. Once again, it was observed that children's first uses of complement clause constructions were highly formulaic and verb-specific. For instance, children's use of verbs such as *think*, *know*, and *see* were restricted to short, formulaic clauses such as *I think + S*, *I know + S, See if + S*. The children in the study gradually began to express propositional content within the main clause as they grew older, although the rate at which children became more productive differed for each child.

The results from the naturalistic research suggest that children learn the syntax of complementation on a verb-by-verb basis, which at face value supports predictions derived from the emergentist position. However, the underlying nature of children's knowledge is not revealed by naturalistic data, and nor is the role that frequency plays in this process. Bloom et al. (1989) only analysed four frequent verbs, and the exact role that frequency plays in Diessel and Tomasello's (2001) data is unclear because those complementizer verbs that children use first and most often are the most frequent in the language. It is therefore possible that their conclusions that children's knowledge of complementation is item-based and not abstract are not warranted because naturalistic data do not reveal children's competence. That is, children (and adults) do not always say everything they know.

We report on an experiment that identifies a role for complementizer verb frequency in children's acquisition and processing of finite sentential complements. Children aged 2;10–5;8 were required to repeat complement clause constructions manipulated for complement-taking verb (CTV) frequency and grammaticality. The CTVs varied in the frequency with which they occurred in a finite sentential complement relative to their overall use in the language. We hypothesised that, if lexical entries are tagged with information about the kinds of syntactic environments they can inhabit, and if this relationship is in some way driven by frequency, children would provide more exact repetitions of sentences containing high frequency (HF) CTVs than those that contain low frequency (LF) CTVs. Based on this same logic, we expected children to make more corrections of ungrammatical sentences containing HF CTVs than those that contain LF CTVs.

1. Method

1.1. Participants

Sixty-one children were recruited from nurseries and primary schools in the Greater Manchester area, and from a registry of volunteer families at the Max Planck Child Study Centre. The sample was divided into two age groups: 'young' children (*n* = 31, mean age: 3;7, range: 2;10–4;2) and

Table 1
Frequency counts for CTVs in study 1

HF CTV	Frequency count	Relative frequency	LF CTV	Frequency count	Relative frequency
Think	365/528	.69	*Pretend*	35/63	.56
Bet	36/51	.71	*Say*	173/474	.28
Hope	17/30	.57	*Know*	78/1668	.05
See	222/3036	.07	*Hear*	0/177	0

'old' children ($n = 30$, mean age: 4;10, range: 4;3–5;9). None possessed any known language impairment or learning difficulty. There were several reasons why we tested children in this age range. First, pilot work showed that few children younger than 2;10 could complete the entire experiment, and many rarely provided full repetitions of any of the test sentences. Second, we were interested in both the nature of children's knowledge of complement clauses and whether they used lexical frequency information when processing this sentence type. It is unclear how abstract children's knowledge of complementation is, even at 3;0; our study was an experimental test of this. Furthermore, if frequency plays a role in structure building we expected to see frequency effects throughout our sample.

1.2. Materials

Thirty-two test sentences (16 sentence pairs) were constructed using eight complement-taking verbs: four high frequency and four low frequency (see Appendix A). To identify the verbs, a corpus study was conducted using data from six children (Adam, Eve, Sarah, Peter, Nina, Naomi) in the CHILDES corpus (MacWhinney, 2000). For each child, the frequency with which each verb occurred in a finite *that*-complement relative to the total number of times that verb occurred in the child's corpus, was computed. The data were then pooled to obtain an overall frequency estimate. The verbs were identified and matched roughly for semantic similarity; the verb pairs chosen were *think-pretend*, *bet-say*, *hope-know*, and *see-hear* (HF verbs first). The frequency counts for each verb are shown in Table 1.

The verb pairs differed in their overall frequency; however, each HF CTV was more frequent than its LF counterpart. In principle it would have been more ideal to have verbs that occur with equal total frequency in corpus count, but differed in the frequency with which they occurred in a complement clause construction. In practice it is very difficult to fulfil this criterion in acquisition research, since the range of verbs that one can use is restricted in comparison to the range that can be used with adults. It was not the case, however, that the HF verb group and LF group verbs differed substantially in their overall base frequency; the materials were as balanced as they could possibly be given the restrictions of acquisition research.

Each verb pair appeared in four sentences manipulating grammaticality: two containing a finite complement clause and the other two containing a non-finite complement clause. These sentences, for the verb pair *think-pretend*, are shown in (1)–(4).

(1) I think/pretend she is riding away on the horse.
(2) I think/pretend he jumps in the swimming pool.
(3) *I think/pretend him running away from the dog.
(4) *I think/pretend her skipping away with her sister.

All test sentences began with the first person singular pronoun *I*, followed by the main verb and sentential complement. Additionally, following the naturalistic data, the main verb was not marked for tense. This format was chosen because Diessel and Tomasello (2001) reported that this form marks children's earliest and most frequent productions of sentential complements. Grammaticality was manipulated by changing the finiteness of the subordinate clause. For example, since *think* and *pretend* take only a finite sentential complement clause, (3) and (4) are ungrammatical because the subordinate clauses in these sentences are non-finite. Note that, since *see* and *hear* also take a non-finite sentential complement, these verbs did not occur in any ungrammatical sentence frames.

Each sentence was preceded by its own discourse context. These were included to ensure a felicitous reading of the test sentences. Sometimes these contexts used some of the same words that occurred in the test sentence. The presence of the same words in the discourse context did not prime children to produce more repetitions of the test sentences: a simple correlation between the number of words contained in the preamble and children's exact repetitions was not significant ($r = .142$, $p = .336$). Toy characters served as referents for the test sentences, providing children with concrete referents onto which they could map the language they heard.

1.3. Procedure

The children were tested individually in a quiet area of their nursery/school. They were told that they were going to play some games with the experimenter. The first game was called the 'parrot game'. In this task the experimenter discussed with the child how a parrot talks by copying 'exactly what people say'. The child was then asked if they could, like a parrot, repeat exactly what the experimenter said. Upon the child's approval, the experimenter read through three example sentences (two simple sentences and a coordinate structure), and asked the child to repeat the sentence after they had heard it, just as a parrot would. If the children made any mistakes the experimenter corrected them, and they were asked to repeat the sentence again. This continued until the child gave an exact repetition. Therefore, the pre-training task served to instruct children that they were required to repeat the sentences read by the experimenter exactly.

The children were then given three more example items that were in the same format as the test items to follow. That is, a discourse context was established which led to one of the toy characters producing an utterance that the child was required to repeat. In particular, it was important that children were aware that they were required to repeat the characters' utterances in their entirety. The training sentences provided an opportunity to correct the children if they did not understand the full requirements of the task.

The test sentences immediately followed the pre-training and training sessions. An example test item proceeded as follows. First, the experimenter told the child which characters were needed for that particular 'story'. The characters were named and put into position on the table in front of a purpose built set. The children were then told the story. An example is given in (5) (the test sentence is underlined):

(5) *Tim and his mum were just about to eat dinner but they couldn't find Dad. 'Maybe he's in the garden', said Mum. 'No', said Tim, <u>'I bet he is talking on the telephone'</u>.*

The experimenter then said to the child 'Can you say that?' The child's response was recorded on a response sheet. Each sentence had its own discourse context. There were two testing sessions. Half of the sentences were presented in the first session, and the other half were presented in the

second session. The sessions were between one and two weeks apart. Each session lasted between 20 and 30 min. Some of the younger children were tested over a greater number of sessions because they did not want to sit through the full sessions. When this occurred the children in question were always re-introduced to the testing procedure using the training items. All sessions were audio-recorded and transcribed to compare to the experimenter's on-line recording. The presentation of test sentences was pseudo-randomized to avoid order effects; eight orders were used.

2. Results

Children's utterances were transcribed and coded for: (i) exact repetitions and, (ii) corrections to ungrammatical sentences. An exact repetition was when the child produced the test sentence in its entirety, word for word. Some minor variation was accepted, such as when children used contraction (e.g., *she's* for *she is*) or when children used acceptable synonyms for words (e.g., *phone* for *telephone*). A number of analyses were conducted. The analyses of children's exact grammatical repetitions are reported, followed by the analyses of children's corrections to the ungrammatical sentences. We then report on children's main verb substitutions.

2.1. Repetitions of grammatical sentences

Table 2 shows the proportion of exact repetitions of grammatical sentences by CTV frequency for each age group.

The data were analysed by conducting two 2 (CTV frequency: High versus Low) by 2 (Age group: Young versus Old) ANOVAs, one by participants (F_1) and one by items (F_2). The dependent measure was the proportion of correct repetitions. Children in both age groups provided significantly more exact repetitions of sentences containing HF CTVs than those that contained LF CTVs, but this was only significant by participants ($F_1(1, 59) = 12.342$, $p = .001$, partial $\eta^2 = .173$, $F_2(1, 28) = 1.805$, $p = .190$, partial $\eta^2 = .061$). This trend held for both age groups (OLD: $t_1(29) = 2.076$, $p = .024$, $t_2(14) = .805$, $p = .217$; YOUNG: $t_1(30) = 2.965$, $p = .003$, $t_2(14) = 1.12$, $p = .141$). The older children provided significantly more exact repetitions of the test sentences than the younger children ($F_1(1, 58) = 14.085$, $p < .001$, partial $\eta^2 = .193$, $F_2(1, 28) = 13.074$, $p = .001$, partial $\eta^2 = .318$). There was no CTV frequency by group interaction.

2.2. Corrections to ungrammatical sentences

An analysis of children's exact repetitions of ungrammatical sentences yielded no significant effects. With respect to children's corrections, we were interested in those corrections where

Table 2
Proportion (and SEs) of exact repetitions of grammatical sentences by age group and CTV frequency

	CTV frequency		
	HF	LF	Total
Young	.339 (.049)	.242 (.046)	.29 (.044)
Old	.567 (.05)	.488 (.047)	.527 (.045)
Total	.453 (.035)	.365 (.033)	.409 (.032)

Table 3
Mean proportion of 'Syntax Only' and 'Syntax + Semantics' corrections and standard errors by age and CTV frequency

	Syntax only		Syntax + Semantics		Total
	HF	LF	HF	LF	
Young	.282 (.044)	.173 (.031)	.105 (.027)	.214 (.035)	.194 (.023)
Old	.379 (.045)	.271 (.031)	.121 (.028)	.225 (.036)	.249 (.023)
Total	.331 (.031)	.222 (.022)	.113 (.019)	.219 (.025)	.221 (.016)

children produced a complement clause construction. These corrections were divided into two categories: (i) corrections that preserved the semantics of the intended message and corrected the syntactic violation only ('syntax only' correction), and (ii) corrections that changed both the syntax and semantics of the sentence ('syntax + semantics' correction). Examples of these corrections for the ungrammatical test sentence (6) are shown in (7) and (8).

(6) *I think him running away from the dog.
(7) I think he's running away from the dog ('syntax only').
(8) I bet he's running from the dog ('syntax + semantics').

Syntax + semantics corrections, such as (8), occurred when children changed the main verb of the original test sentence. These corrections are interesting because the kinds of verbs children substitute into their repetitions give an indication of the verbs that are best attested in this structure. The number of 'syntax only' and 'syntax + semantics' corrections, as a proportion of the total number of ungrammatical sentences, are shown in Table 3.

Table 3 shows that, overall, children made more 'syntax only' than 'syntax + semantics' corrections, but that correction type appeared to interact with CTV frequency: children in both age groups made more 'syntax only' corrections when the test sentence contained a HF CTV, but made more 'syntax + semantics' corrections when the test sentence contained a low frequency CTV. Each correction type was analyzed separately.

The syntax only corrections were analyzed first. Two 2 (CTV frequency: High versus Low) by 2 (Age group: Young versus Old) ANOVAs were conducted, one by participants and one by items. The proportion of 'syntax only' corrections was the dependent measure. Children in both age groups made significantly more 'syntax only' corrections when the test sentence contained a HF CTV ($F_1(1, 59) = 17.528$, $p < .001$, partial $\eta^2 = .229$, $F_2(1, 20) = 3.925$, $p = .061$, partial $\eta^2 = .164$). This was true for both age groups (OLD: $t_1(29) = 3.209$, $p = .0015$, $t_2(10) = 1.514$, $p = .081$, YOUNG: $t_2(30) = 2.777$, $p = .0045$, $t_2(10) = 1.28$, $p = .115$). The older age group made significantly more 'syntax only' corrections than the younger group ($F_1(1, 59) = 4.167$, $p = .046$, partial $\eta^2 = .066$, $F_2(1, 20) = 3.914$, $p = .062$, partial $\eta^2 = .164$). The frequency by group interaction was not significant.

The 'syntax + semantics' corrections were analyzed in the same manner using the proportion of these corrections as the dependent measure. Children in both groups made significantly more 'syntax + semantics' corrections when the test sentence contained a LF CTV ($F_1(1, 59) = 24.271$, $p < .001$, partial $\eta^2 = .291$, $F_2(1, 20) = 3.157$, $p = .091$, partial $\eta^2 = .136$). This was true for both age groups (OLD: $t_1(29) = 4.334$, $p < .001$, $t(10) = 1.16$, $p = .132$, YOUNG: $t_1(30) = 3.057$, $p = .003$, $t_2(10) = 1.348$, $p = .104$). No other effects were significant.

Table 4
Number of times each verb was substituted into children's recall

Verb	Frequency	Age	
		Young	Old
Think	High	179	108
Hope	High	24	48
Bet	High	34	28
See	High	11	14
Total		248	198
Pretend	Low	1	1
Know	Low	9	14
Say	Low	4	12
Hear	Low	4	4
Total		18	31

2.3. Main verb substitutions

Overall, children substituted a new main verb into their repetition of the test sentence on 25.37% of trials. These including trials where children substituted a new main verb to correct an ungrammatical test sentence, and trials where they substituted a new main verb into grammatical test sentences. Table 4 shows the number of times each verb was substituted into children's repetitions.

Children in both age groups substituted more HF CTVs than LF CTVs into their repetitions. The data were analyzed separately for each age group using single sample proportion tests. The dependent variable was a number of HF CTV substitutions divided by the total number of substitutions (i.e., HF substitutions/(HF + LF substitutions). This was computed for each child. The proportion was tested against a baseline proportion of .5; that is, the null hypothesis was that the number of substitutions for both frequency categories was equal. Both age groups substituted significantly more HF CTVs than LF CTVs into their repetitions (YOUNG: Sample proportion: .932, $z = 17.035$, $p < .001$, CI_{95}: [.899, .959]; OLD: Sample proportion: .865, $z = 12.367$, $p < .001$, CI_{95}: [.817, 906]).

This suggests that children more often 'traded up' to HF CTVs both when correcting ungrammatical sentences and when recalling test sentences. The verb that was substituted with the highest frequency was *think*. In Diessel and Tomasello's (2001) data, *think* was the most common CTV to occur with a *that*-complement, and was the most frequent CTV found in the children's input. This fact, along with the pervasiveness of *think* in children's repetitions, suggests there is something special about this verb to children's acquisition of finite sentential complements. We consider this finding further in Section 3.

2.4. Analysis by raw frequency

It is possible that the results of our study could be explained by raw frequency. That is, rather than the relative frequency with which a CTV occurs in a sentential complement construction, the frequency effect could be due to the encoding and retrieval of lexical items (i.e., CTVs that are more frequent overall are easier to retrieve from memory). This is a less interesting explanation, since it would mean that we would not have a basis to argue that individual CTVs are tagged with

information about their distributional co-occurrence. In turn, this would also argue against a role for lexical frequency in the entrenchment process. In order to test this alternative explanation, we re-ran our analyses using raw frequency as an independent variable. With respect to children's exact repetitions, there was no significant main effect for frequency ($F_1(1, 59) = .415$, $p = .522$, partial $\eta^2 = .007$, $F_2(1, 28) = .049$, $p = .827$, partial $\eta^2 = .002$), and nor was there a significant frequency by group interaction ($F_1(1, 59) = 1.805$, $p = .184$, partial $\eta^2 = .03$, $F_2(1, 28) = .212$, $p = .649$, partial $\eta^2 = .008$).

There were no frequency effects found in children's corrections either. For the syntax only corrections the main effect for frequency was not significant ($F_1(1, 59) = .675$, $p = .415$, partial $\eta^2 = .011$, $F_2(1, 20) = .12$, $p = .733$, partial $\eta^2 = .006$) and nor was the frequency by group interaction ($F_1(1, 59) = .006$, $p = .938$, partial $\eta^2 < .001$, $F_2(1, 20) = .001$, $p = .974$, partial $\eta^2 < .001$). For the syntax + semantics corrections the main effect was not significant ($F_1(1, 59) < .001$, $p = .995$, partial $\eta^2 < .001$, $F_2(1, 20) < .001$, $p = .997$, partial $\eta^2 < .001$) and nor was the interaction ($F_1(1, 59) = .176$, $p = .676$, partial $\eta^2 = .003$, $F_2(1, 20) = .043$, $p = .838$, partial $\eta^2 = .002$). On the basis of these results we can conclude that raw CTV frequency cannot account for children's ability to repeat and correct sentential complement constructions.

3. Discussion

The results suggest that the relative frequency with which a complementizer verb occurs in a complement clause construction predicts children's ability to remember and repeat that construction. Both of our predictions were upheld: children produced more exact repetitions of sentences containing HF CTVs, and were more likely to correct an ungrammatical sentence using a HF CTV. There was no frequency by age interaction; both age groups were sensitive to the frequency manipulation. These results support an account of language acquisition where lexical frequency information plays an important role in both the acquisition of syntactic knowledge and the processing of syntactic structure.

With respect to children's syntactic processing, the results identify a role for lexical frequency in structure building by children, and therefore identify continuity between the developing parser and the adult state, since similar results have been reported for adult participants (see Garnsey et al., 1997; Holmes et al., 1989; Trueswell et al., 1993). The results therefore suggest that from an early age children keep track of the distributional environments in which individual verbs appear, and annotate this frequency information onto their lexical entries for verbs. Such results sit well with constraint-satisfaction approaches to processing (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Seidenberg & MacDonald, 1999), which argue that the parser pursues the interpretation that is best attested in the input.

The results also sit well with the emergentist approach to acquisition, which argues that verb frequency is important in the entrenchment process; children learn lexico-constructional pairings that become gradually abstract with experience. It appears that even our youngest children possessed some abstract knowledge of complementation. This is clear from children's substitutions, which suggest that children are able to separate form from meaning—the hallmark of abstraction. This conclusion stands somewhat in contrast to the conclusions drawn from studies of children's naturalistic speech (Bloom et al., 1989; Diessel & Tomasello, 2001), although it does not explicitly argue against the mechanisms by which these researchers argue children acquire language.

The children's main verb substitutions provide a potentially unique insight into their knowledge of complementation and the process by which they generalize over their input. The children overwhelmingly substituted *think* into their repetitions, suggesting that there is something special

about this verb in this syntactic frame. Goldberg and co-workers (Casenhiser & Goldberg, 2005; Goldberg, 1999, 2005; Goldberg et al., 2004) have argued that semantically general, high frequency verbs which occur with high frequency in individual constructions facilitate the acquisition of syntax. They suggest that a single high frequency exemplar has a facilitatory effect in learning argument structure generalizations because the exemplar provides a rough estimation of the syntax-semantics correspondence for the construction type. For example, Goldberg et al. (2004) analyzed children's and mothers' use of three construction types (intransitive motion, caused motion, dative) and showed that one verb accounted for the 'lions share' of tokens for each construction in both the children's and mothers' speech. For instance, for the dative the most frequent verb was *give*. The verbs are semantically general; Goldberg et al. argued that these verbs underlie the constructional meaning, which in the case of the dative is object transfer. Therefore, these high frequency verbs are thought to aid children in making argument structure generalisations (for experimental support see Casenhiser & Goldberg, 2005).

It is possible that *think* serves a similar function for finite sentential complements, and that the pervasiveness of *think* in children's repetitions may represent a vestige of early exemplar-based learning. In terms of frequency, *think* had privileged status in the child corpora (Table 1) because it occurred highly frequently with a sentential complement and was highly frequent overall. Both Thompson & Mulac (1991) and Roland, Elman, and Ferreira (2006) observed similar results in studies examining adult corpora. What kinds of information would the exemplar NP + *think* + finite S give children? Following Goldberg et al. (2004), we suggest that it could provide them with a wealth of information about the syntax of complementation. First, as argued by Goldberg et al., a verb-based exemplar would give an initial approximation of the syntax-semantics correspondences for this construction type. The information gleaned from an exemplar could serve as a basis to learn more complementizer verbs and more facts about complementation. Following arguments made by Gleitman and co-workers (Fisher, 1996; Gleitman, 1990; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005), children could exploit the structural pattern provided by the exemplar to learn new CTVs; that is, they could acquire new verbs via syntactic bootstrapping. Once children had learned a critical mass of CTVs they could also exploit the semantics of verbs to infer that new verbs belonging to the same semantic class behave similarly syntactically, as proposed by Pinker (1989). From here children can acquire more aspects of the language using multiple evidential sources of information; that is, both syntax and semantics (Seidenberg & MacDonald, 1999).

This potential explanation for the data is consistent with lexicalist proposals of acquisition and processing (Bates & MacWhinney, 1989; Seidenberg & MacDonald, 1999). According to these approaches, no sharp distinction is drawn between knowledge of language and processing of language; they are two sides of the same coin. It must be stressed that this is an idea that is far from universally accepted. The relationship between processing and acquisition is yet to be elucidated to a sufficient extent that we can confidently argue for collapsing the two concepts. However, we argue that the results from the present study suggest that this is a hypothesis worth pursuing.

We have interpreted the results of the present study to suggest that lexical frequency information aids structure building; that is, that structure is predicted given verb frequency information (i.e., Pr(construction|verb). This is consistent with the general emphasis in linguistics on the verb as a key determinant of sentence meaning (e.g., Grimshaw, 1990). However, the fact that children made many main verb substitutions suggests that the information available from the complement clause construction itself is important. That is, the children's main verb substitutions suggest that when they lost a verbatim representation of the sentence they were computing important

lexical content on the basis of the argument structure frame (i.e., Pr(verb|construction). This provides experimental support for recent arguments made by Goldberg, Casenhiser, and Sethuraman (2005), who used corpus data to argue for the importance of constructions to sentence meaning. Verbs can occur in multiple syntactic environments and encode different meanings depending on the construction in which they occur. Therefore, verbs are not always as reliable a cue to sentence meaning as are constructions. This is because a smaller range of constructions are available to encode a given message; the pool from which to express, for instance, descriptions of mental states (as in complement clause constructions), is necessarily smaller than the pool of verbs that can describe individual mental states. The suggestion is that attention to sentence-level constructions is essential for children to make argument structure generalizations. The data from the present study show, minimally, that the children in our study treated verb and construction as semi-independent entities, and that each was used in the process of reconstructing the test sentence during repetition.

In this paper we have demonstrated that lexical frequency information plays an important role in children's processing and acquisition of finite complement clause constructions. More research is required to flesh out our arguments. In particular, converging cross-linguistic evidence from languages that are typologically different from English is required. The emergentist position makes specific predictions that are language-specific. For instance, as argued by theorists such as Bates and MacWhinney (1989), children will acquire their input language in accordance with the concepts of *cue reliability*; children become sensitive to those aspects of their input language that provide the most 'payoff'. We stress that an accurate and detailed analysis of the input must be conducted in order for cue reliability estimates to be meaningful.

Lexical frequency will not, of course, be the sole determinant of acquisition. For instance, conceptual complexity is likely to affect the acquisition of individual complementizer verbs; children must be able to represent and understand the concept a verb encodes in order to use it, which in the case of complement clauses means being able to represent mental states (e.g., see Astington & Jenkins, 1999; Bartsch & Wellman, 1995). Our argument, therefore, is not that lexical frequency is the only determinant of language acquisition. Rather, we argue that frequency is one, albeit important, source of information. This paper goes some way to increasing our understanding of how this information could be important.

Acknowledgements

Appendix A

Think/pretend

(1) I think/pretend she is riding away on the horse.

   Alice and her mum and dad were looking at their horse. Alice decided to hop on the horse and go for a ride. 'Where's Alice going', asked Dad. Mum said, '*I think she is riding away on the horse*'.

Alice and Jenny were playing with their dolls. Jenny asked Alice what she liked to do most when she was playing with her favourite doll, Alice said '*I pretend she is riding away on a horse*'.

(2) I think/pretend he jumps in the swimming pool.

Johnny and Tim were watching a film. Johnny had seen the film before, and at a really exciting part Tim asked him what happens next. Tim said, '*I think he jumps in the swimming pool*'.

Johnny and Tim were making up stories about their pretend friends. Johnny was telling Tim a story about his friend Peter. 'What happens next?' Tim asked. '*I pretend he jumps in the swimming pool*'.

(3) *I think/pretend him running away from the dog.

Johnny and Tim were playing in the street where the big dog lived. Johnny was scared of the dog, and ran off while Tim wasn't looking. Johnny's mum was looking for him. 'Where's Johnny', she asked Tim. '*I think him running away from the dog*'

Johnny and Tim were making up stories again. Johnny was telling a story about a dog and a little boy. Johnny said, '*I pretend him running away from the dog*'.

(4) *I think/pretend her skipping away with her sister.

Alice and Jenny were playing in the garden with Jenny's sister. Jenny and her sister decided to skip around the block, but didn't tell Alice. Here's Alice's mother. 'Where's Jenny?', she asked. '*I think her skip away with her sister*'.

Alice and Jenny were making up stories. Alice was telling a story about her friend Sarah and her sister. She said, '*I pretend her skip away with her sister*'.

Hope/Know

(5) I hope/know she is making a chocolate cake.

Alice was sitting with her dad, and her mum was in the kitchen. 'What's mum doing?', asked dad. Alice said, '*I hope she is making a chocolate cake*'.

Alice was sitting with her dad, and her mum was in the kitchen because Alice asked her to make a chocolate cake. 'Mum's making a chocolate cake', said dad. Alice said, '*I know she is making a chocolate cake*'.

(6) I hope/know he plants some pretty flowers.

Tim and his mum were watching dad work in the garden. Dad was planting lots of things. Tim said, '*I hope he plants some pretty flowers*'.

Tim and his mum were talking about the gardener who lives across the road. 'What do you know about the gardener?', asked mum. Tim said, '*I know he plants some pretty flowers*'.

(7) *I hope/know her driving to the toy store.

Tim's mum was driving Tim and Johnny to the toy shop. 'Mum's driving to the toy shop', said Tim. Johnny said, '*I know her driving to the toy shop*'.

Tim and Johnny were in the car and Tim's mum was driving. Tim said, '*I hope her driving to the toy shop*'.

(8) *I hope/know him buying the yummy sweets.

Alice and Jenny were in the supermarket with Alice's dad. Alice and Jenny both like sweets. Alice whispered to Jenny, '*I hope him buy some yummy sweets*'.

Tim and Johnny were in the in sweets shop with Tim's mum. 'Johnny is going to buy some sweets', said Tim's Mum. '*I know him buy some yummy sweets*', said Tim.

Bet/Say

(9) I bet/say he is talking on the telephone.

Tim and his mum were just about to eat dinner but they couldn't find dad. 'Maybe he's in the garden', said mum. 'No', said Tim, '*I bet he is talking on the telephone*'.

Alice and Jenny were wondering where Jenny's dad had gone. 'Maybe he's sleeping in his room', said Alice. 'No', said Jenny, '*I say he is talking on the telephone*'.

(10) I bet/say she eats chocolate ice-cream.

Alice and Jenny were watching a fun video about a lady eating yummy food. 'Maybe she will eat some sweets now', said Alice. 'No', said Jenny, '*I bet she eats some chocolate ice-cream*'.

Alice and Jenny were playing a game with their dolls. 'What will the dolly do next?' asked Alice. '*I say she eats some chocolate ice-cream*', said Alice.

(11) *I bet/say him jumping over the fence.

Johnny and Tim were watching a film where a man was riding a fast motorbike. 'He might crash into the fence', said Johnny. 'No', said Tim, '*I bet him jumping over the fence*'.

Johnny and Tim were playing a game with their action men. 'What happens next?', asked Johnny. Tim said '*I say him jumping over the fence*'.

(12) *I bet/say her give the present to her mum.

Alice and Jenny went to a birthday party for their friend Sarah. Alice gave Sarah a present but she couldn't see where Sarah put it. 'Maybe she lost it?', said Jenny. 'No', said Alice, '*I bet her give the present to her mum*.

Alice and Jenny were shopping for presents with Jenny's mum. They bought lots of presents for all their family. Jenny's mum asked Jenny, 'Who will Alice give that present to?' Jenny said '*I say her give the present to her mum*'.

See/Hear

(13) I see/hear he is going to the movies.

Johnny was going to meet Tim at the cinema. Johnny's dad didn't know where Johnny was going but saw him leave. He looked out the window and said, '*I see he is going to the cinema*'.

Johnny told Tim he was going to the cinema for his birthday. After Johnny had left, Tim's mum asked him about Johnny's birthday. '*I hear he is going to the cinema*', she said.

(14) I see/hear she cooks dinner for the family.

Johnny and Tim were on holidays at Tim's auntie's house. It was dinner time and Tim's cousin, Sarah, was cooking dinner. Johnny said '*I see she cooks dinner for the whole family*'.

Alice and Jenny were talking about their friend Janet. 'Janet loves cooking', said Jenny. 'That's right', said Alice, '*I hear she cooks dinner for the whole family*'.

(15) I see/hear him running up the stairs.

Johnny and Tim were being naughty at school. Their teacher, Mr. Wood, stopped Johnny in the corridor and said, 'where's Tim?'. Johnny looked up and said, '*I see him running up the stairs*'.

Johnny and Tim were playing a game of hide-and-seek. Tim was trying to find Johnny. Tim heard a noise above him. 'Oh', he said, '*I hear him running up the stairs*'.

(16) I see/hear her sing the song about the blind mice.

Alice and her Mum went to the theatre. When they got home her daddy asked her what she saw. Alice said she saw a lady singing. 'What did you see her singing?', asked her daddy. Alice said, '*I see her sing the song about the blind mice*'.

Alice and Jenny were at home playing with their dolls. 'I wonder where mummy is', said Alice. 'Oh, that's funny', said Jenny, '*I hear her sing the song about the blind mice!*'

References

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind. *Developmental Psychology*, *35*, 1311–1320.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York: Oxford University Press.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–73). Cambridge: Cambridge University Press.

Bloom, L., Rispoli, M., Gartner, B., & Hafitz, J. (1989). Acquisition of complementation. *Journal of Child Language*, *16*, 101–120.

Brooks, P. J., Tomasello, M., Dodson, K., & Lewis, L. B. (1999). Young children's overgeneralizations with fixed transitivity verbs. *Child Development*, *70*, 1325–1337.

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science*, *8*, 500–508.

Diessel, H. (2004). *The acquisition of complex sentences*. Cambridge, UK: Cambridge University Press.

Diessel, H., & Tomasello, M. (2001). The acquisition of finite complement clauses in English: A corpus-based analysis. *Cognitive Linguistics*, *12*, 97–141.

Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, *31*, 41–81.

Garnsey, S., Pearlmutter, M., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*, 58–93.

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

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

Goldberg, A. E. (1999). The emergence of argument structure semantics. In B. MacWhinney (Ed.), *The emergence of language* (pp. 197–212). Mawah, NJ: Erlbaum.

Goldberg, A. E. (2005). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalisations. *Cognitive Linguistics*, *15*, 289–316.

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2005). The role of prediction in construction-learning. *Journal of Child Language*, *32*, 407–426.

Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.

Holmes, V. M., Stowe, L., & Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language*, *28*, 668–689.

Lapata, M., Keller, F., & Schulte im Walde, S. (2001). Verb frame frequency as a predictor of verb bias. *Journal of Psycholinguistic Research*, *30*, 419–435.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

MacWhinney, B. (1998). Models of the emergence of language. *Annual Review of Psychology*, *49*, 199–227.

MacWhinney, B. (Ed.). (2000). *The CHILDES Project: Tools for analysing talk: Vol. 1*. Mawah, NJ: Erlbaum.

Matthews, D., Lieven, E. V. M., Theakston, A. L., & Tomasello, M. (2005). The role of frequency in the acquisition of English word order. *Cognitive Development*, *20*, 121–136.

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

Roland, D., Elman, J., & Ferreira, V. (2006). Why is that? Structural prediction and ambiguity resolution in a very large corpus of English sentences. *Cognition*, *98*, 245–272.

Seidenberg, M., & MacDonald, M. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, *23*, 569–588.

Theakston, A. L. (2004). The role of entrenchment in children's and adult's performance on grammaticality judgment tasks. *Cognitive Development*, *19*, 15–34.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2004). Semantic generality, input frequency and the acquisition of syntax. *Journal of Child Language*, *31*, 61–99.

Thompson, S. A., & Mulac, A. (1991). The discourse conditions for the use of the complementizer 'that' in conversational English. *Journal of Pragmatics*, *15*, 237–251.

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

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden- paths. *Journal of Experimental Psychology: Learning, memory and Cognition*, *19*, 528–553.