This is a contribution from *Acquisition and Development of Hebrew. From infancy to adolescence*.
Edited by Ruth A. Berman.

The nature of CDS in Hebrew

Frequent frames in a morphologically rich language

Inbal Arnon
Hebrew University Jerusalem

The chapter explores the distribution and content of frequent frames – recurring multiword sequences – appearing at the start of utterances in speech directed to young Hebrew-speaking children. Previous work has documented the existence and prevalence of such frames in several languages (English, German, and Russian). Here, analysis of a dense corpus of Hebrew child-directed speech was undertaken with two goals in mind. The first aims at examining the *distribution* of multiword elements in Hebrew child-directed speech, to ascertain whether frequent frames are found in a morphologically rich language like Hebrew and, if so, to see how pervasive they are compared to other languages, and how consistent across different caretakers. The second goal is to explore the *content* of frequent frames in Hebrew to address such questions as: Do they provide children with relevant morphological and syntactic information? Are they frequent enough to be employed in learning? Results show that frequent frames do occur in Hebrew, that they are relatively consistent across caretakers, and that they illustrate a range of grammatical relations. These findings expand our understanding of frequent frames in general, while also adding to the relatively sparse information on the nature of child-directed speech in Hebrew.

Keywords: child-directed speech; corpus study; cross-linguistic; distributional information; frequent frames; gender-marking; grammatical relations; Hebrew; inflection; input; language acquisition; learning; morphology; morphological acquisition; multiword units; syntactic acquisition; usage-based; variation

Introduction

Words are often seen as the basic building blocks of language representation and use: the units from which larger sequences and sentences are created. The distinction between words, on the one hand, and multiword combinations, on the other, has a long history in the study of language and is central to linguistic theories that assume a clear division between lexicon and grammar (Chomsky 1965; Pinker 1999). This

DOI 10.1075/tilar.19.07arn

distinction has also impacted language acquisition research. As Ann Peters eloquently argued several decades ago, our characterization of the linguistic units that adults utilize in language shapes the questions we ask, and the units we look for, in child development. If words are seen as basic building blocks but larger units are not, the researcher will look for, and focus on, children's acquisition of words rather than larger sequences (Peters 1977, 1983). Textbook descriptions of language learning implicitly reinforce this distinction: Children's progression is often characterized as a move from smaller building blocks to larger combinations, from syllables via words to multiword combinations and complex constructions. This description assumes that words and multiword sequences are learned in a serial fashion: Children first learn words and only then how to combine them.

Recent years have seen a shift away from this perspective. There is growing theoretical emphasis on – and empirical evidence for – the idea that multiword units, like words, are basic building blocks in language learning and use. This idea is found in linguistic approaches that emphasize the role of *constructions* in language (Culicover & Jackendoff 2005; Goldberg 2006: 290; Langacker 1987: 540) and is advocated in single-system models of language which assume that all linguistic material – whether a word or a sequence – is processed by the same cognitive mechanisms (Bybee 1998; Elman 2009; McClelland 2010). In support of these models, there is growing evidence for parallels in processing between words and larger sequences: Speakers have been shown to be sensitive to the distributional properties of multiword sequences in production and comprehension (e.g., Arnon & Cohen Priva 2013, 2014; Arnon & Snider 2010; Reali & Christiansen 2007; Tremblay, Derwing, Libben, & Westbury 2011). Adults, for example, are faster to respond to higher-frequency phrases (Arnon & Snider 2010) while young children are faster and more accurate at producing higher-frequency phrases (Arnon & Clark 2011; Bannard & Matthews 2008).

From a developmental perspective, the distinction between words and larger units is challenged by usage-based approaches to language learning (e.g., Bannard, Lieven, & Tomasello 2009; Lieven & Tomasello 2008; Tomasello 2003). In such approaches, grammatical knowledge is learned by abstracting over stored exemplars of varying sizes and levels of abstraction. This process is expected to be highly sensitive to usage patterns, a prediction that is borne out in numerous studies (for reviews, see Diessel 2007; Lieven 2010). Multiword units play an important role in such models (Abbot-Smith & Tomasello 2006; Arnon 2010; Bod 2009). On the one hand, they provide children with lexically specific chunks to use in early production, so serving as building blocks for early language. At the same time, they allow children to discover grammatical relations that hold between words, so constituting units that children utilize in learning. Consequently, there is increased interest in the content and nature of multiword information in child-directed speech. If children attend to and learn from multiword information, then (a) such information should be consistently found in

child-directed speech, and (b) it should provide children with grammatically relevant information about the structural relations they need to learn.

This chapter explores the multiword information available to Hebrew-speaking children, focusing on *frequent-frames*, that is, recurring multiword sequences appearing in child-directed speech.[1] Previous work documenting the existence of such frames in child-directed speech has shown them to be prevalent and relatively consistent across caretakers (Cameron-Faulkner, Lieven & Tomasello 2003). Follow-up studies indicate that such frames are also found in languages with a richer morphology than that of English, although there are interesting cross-linguistic differences in their nature and distribution (Stoll, Abbot-Smith & Lieven 2009). Here, I look at frequent frames in speech directed to young Hebrew speakers with two goals in mind. The first goal is to examine the *distribution* of multiword information in Hebrew child-directed speech; in particular, to ascertain whether frequent frames are found in a morphologically rich language like Hebrew and, if so, to see how pervasive they are (compared to other languages) and how consistent across different caretakers. A second goal is to explore the *content* of frequent frames in Hebrew: Do they provide children with morphological and syntactic information? Are they frequent enough to be used in learning? Studying these questions should expand our understanding of frequent frames, as well as the nature of child-directed speech in Hebrew. The following sections provide background on the study of frequent frames and review relevant structural properties of Hebrew.

Frequent frames in child-directed speech

Cameron-Faulkner and colleagues were the first to systematically study recurring multiword sequences in child-directed speech (Cameron-Faulkner et al. 2003). Up to that point, most studies of child-directed speech focused on the relation between child language and the distribution of specific items such as nouns or verbs (e.g., Naigles & Hoff-Ginsburg 1998) or particular syntactic constructions (e.g., Huttenlocher, Vasilyeva, Cymerman & Levine 2002) in the input addressed to children. The goal of Cameron-Faulkner and her associates was, rather, to characterize the full range of utterances that children are exposed to, and to see if these contain recurring multiword patterns that might facilitate language learning. To attain this goal, they

1. The term *frequent-frame* refers to two different notions in the developmental literature. In the usage-based tradition, it is used to describe recurring sequences; while the term is not limited to utterance-initial position, in practice, research has focused on utterance-initial sequences. In the syntactic category learning literature (Mintz 2003) – as reviewed further in the general discussion – the term refers to a distributional pattern defined as two words that bracket one intervening word (e.g., *I __ you*), with such frames considered to assist grammatical category learning.

used a dense sample of adult-child interactions in the so-called 'Manchester corpus' (Theakston et al. 2001) to analyze utterance-initial sequences produced by 12 English-speaking mothers. They choose to focus on utterance-initial rather than utterance-medial sequences for both practical and theoretical reasons: Such sequences are easier to extract from a corpus, they have a clear starting point and share certain utterance-initial prosodic features, and they may be more perceptually salient compared to sequences positioned in the middle of utterances (Seidl & Johnson 2006).

In the seminal study of Cameron et al., utterances were first coded for construction type (questions, imperatives, routines, etc.), and their initial multiword sequences were then identified and analyzed to see if they qualify as frequent frames. In this context, a *frequent frame* was defined as a sequence that appeared at least four times in the speech of a particular mother (across the entire corpus for that mother), while *core-frames* were ones that appeared in the speech of more than half of the mothers. Examples of frequent-frames included questions like 'Are you-', Copula constructions like 'That is-', and complex constructions like 'I think-'. This analysis revealed that children hear recurring frames; that such frames are consistent across mothers, and that they account for a high proportion of the utterances directed to the children in the sample. Over half of maternal utterances were accounted for by 52 frequent-frames. Moreover, children used many of the frequent frames in their own speech, at a rate that was correlated with maternal usage.

These findings raise interesting questions about cross-linguistic variation. If frequent frames play a role in language learning, we would expect to find them across languages. However, the propensity of frames in English may reflect language-specific properties: Utterance-initial sequences are more likely to be repeated in a language with rigid word order and relatively impoverished morphology, like English. For example, the sequence 'are you' is the same in English regardless of the gender or number of the interlocutor. In Hebrew, the form of both the pronoun and the verb would vary according to the gender and number of participants, resulting in fewer repetitions of the exact same sequence.

The only study to examine cross-linguistic variation in the distribution of frequent frames known to me is the comparison by Stoll et al. (2009) of frequent-frames in English, German, and Russian, three typologically different languages. Using coding and extraction methods similar to those of Cameron-Faulkner et al. (2003), they found that frequent frames exist in all three languages, but that their rate, length, and coverage (the proportion of utterances they accounted for) was impacted by language-specific characteristics. Russian had fewer sentence-initial frames than the other two languages, and they accounted for less of the data compared to English and German. This is expected, given Russian's more flexible word order and richer inflectional system. Importantly, frequent frames still accounted for around 70% of child-directed utterances even in Russian. Russian also had very few three-word frames compared

to German and English, probably due to a lack of present-tense copulas and rich inflectional morphology. For example, copula constructions like 'this is' are classified as frequent-frames in English, but the equivalent construction in Russian will only be one-word long (no copula) and will include gender and number marking on the demonstrative, so reducing its frequency. Examination of these three languages shows that frequent frames do appear cross-linguistically, but that they vary in line with language-specific properties.

Relevant structural properties of Hebrew

Hebrew has several structural features that make it an interesting case study for looking at the distribution of frequent frames cross-linguistically. In terms of word order variation, it lies somewhere between English and German. While Hebrew is basically an SVO language (Berman 1978, 1985, 1990), word order is freer than in English and OVS configurations are not infrequent, although they are pragmatically marked (Givón 1973, 1976). However, Hebrew observes stricter word order compared to German – where there are several permissible orders – and Russian – which has highly flexible word order. If word order variability impacts the distribution of frequent frames, we would expect Hebrew to have fewer frames than English, but more than German and Russian.

Hebrew has several morphological features that may also impact the distribution of frequent frames. On the one hand, Hebrew is a morphologically rich language where most parts of speech (verbs, nouns, adjectives, pronouns) are specified for number and gender. In addition, Hebrew verbs, in main and embedded clauses, agree with the subject in number, gender, and often person (agreement is marked for all persons in past and future, but not in the present tense – see Lustigman; Kupersmitt & Meir, this volume). These features will lead to a reduction in repeated sequences compared to languages with less morphological marking, since the same utterance may have many different forms, depending on the gender, number, and person of participants. In this respect, Hebrew is more similar to Russian and German than to English. At the same time, Hebrew has other morpho-syntactic features that may increase the number of frequent frames relative to German and Russian. First, unlike German, Hebrew does not have a fully productive case-marking system, nor does it mark case or gender on articles, so that that the same utterance-initial form can be used in a wider range of situations. Second, unlike Russian, Hebrew does have articles, but does not have case marking on question words, so that the same sequence is likely to recur more in utterance-initial position. From a morphological perspective, then, Hebrew can be expected to have more frequent frames than Russian and German.

In sum, the current study has two goals. The first is to examine the distribution of frequent frames in Hebrew to ask (a) if there are frequent frames in Hebrew, and

(b) what is the effect of language-specific features on their distribution. The second goal is to look at the content of frequent frames in Hebrew to see what kinds of syntactic and morphological information they provide.

Methods and materials

Data

The data for this study was taken from the BSF-Long corpus, a longitudinal corpus of four Hebrew-speaking children collected, and transcribed in broad phonemic transcription in Ruth Berman's lab at Tel Aviv University in the 1990s (as reported, for example, in: Armon-Lotem & Berman 2003; Berman 1990; Berman & Lustigman 2014; Berman & Ravid 2000; Nir 2010; and Uziel-Karl 2001, 2006; as well as in Lustigman, this volume). This same database was subsequently re-transcribed to accord with Hebrew orthography and further analyzed in cooperation with Brian MacWhinney of Carnegie-Mellon University (Albert, MacWhinney, Nir & Wintner 2013).

This corpus was chosen because it is most similar in both density and size to the corpora used in studies of frequent frames in other languages. The database consists of naturalistic longitudinal recordings collected on a weekly basis from four Hebrew-speaking children, one boy (Leor) and three girls (Hagar, Lior, Smadar). All children were monolingual Hebrew speakers raised in upper-to-middle class families with highly educated parents. The children were recorded weekly at home, in the presence of a primary caregiver, for periods ranging from one to three years. The database is coded in broad phonemic transcription with detailed contextual information. Crucially for our purposes, the database contains data from several different children who were all recorded weekly in similar naturalistic situations. This enables us to compare frequent frames produced by different caretakers in relatively similar situations.

In line with previous studies, analysis was confined to speech produced by mothers (rather than other adult caretakers), and directed to the children themselves up until age 2;6 years. Recordings were analyzed only for the three girls, for several reasons: First, the boy Leor was recorded by his maternal aunt, not by his mother like the three girls and, second, his sample yielded fewer adult-child interactions than those of the three girls. It was important, too, to reduce the variability introduced by the child's gender: Since Hebrew marks gender on pronouns and verbs, forms directed to boys often differ from those directed to girls (e.g., *ma at roca* 'what you-SG.FM want-PRS.SG.FM' versus *ma ata roce* 'what you-SG.MS want-PRS.SG.MS', both meaning 'What do you want?'). While this variability may impact the developmental trajectory in interesting ways, data would be needed from more than one boy in order to dissociate the effect of caretaker from that of the child's gender, since different caretakers may use frames differently.

The CLAN program (MacWhinney 2000: 418) was used to extract all child-directed utterances produced by the mothers, yielding a total of 25,716 child-directed utterances across the entire corpus (excluding unintelligible strings), more than in the largest English-based study analyzed to date (Cameron-Faulkner et al. 2003). The utterances were on average 4.3 words long, with 1.7 utterances per turn. The corpora for the three children differed in size, as follows: Hagar's corpus was the largest (16,222 utterances), followed by Lior (5,375 utterances) and Smadar (3,362 utterances). Importantly, all three corpora are larger than those used in previous studies to study frequent frames in morphologically rich languages (Stoll et al. 2009 used 1,400 utterances for each language). The implications of these size differences, and the possible differences between the children are discussed in the results section.

Coding

To meet the goal of investigating the nature and distribution of frequent frames in speech directed to Hebrew-speaking children, all the lexical strings occurring at the beginning of utterances were analyzed. The study included only turn-initial utterances, which can be easily identified automatically. As in previous studies, (Cameron-Faulkner et al. 2003; Stoll et al. 2009), the focus was on utterance-initial frames for practical reasons: such frames were the ones analyzed in previous studies, they are easier to identify in a corpus, and may be more perceptually salient for children. CLAN programs were used to automatically extract all the three-word sequences appearing at the start of child-directed utterances (after excluding unintelligible utterances). The coding conventions in the Hebrew corpus made it possible to treat the seven functors that are affixed as part of the next word in written Hebrew (the definite article *ha-* 'the', prepositions *be-* 'in, at', *le-* 'to', *mi-* 'from', and *ke-* 'as, like', and the conjunctions *ve-* 'and' and *še-* 'that') as separate words, in order to render word counts consistent with the European languages studied by Stoll et al. (see Berman 2002). For example, the sequence *ba-báyit* 'in (the) house' was treated as two words and not one. The frequency of occurrence for each three-word sequence was automatically computed using the co-occur command in CLAN.

Following Cameron-Faulkner et al. (2003), a frequent frame was defined as one appearing at least four times in the speech of each mother (across the entire corpus, excluding immediate repetitions). True, what counts as frequent is of course relative to the size of the corpus, yet this numerical criterion has been used in the past for analyzing corpora of different sizes, including ones smaller than our smallest sample (1,400 utterances in Stoll et al. 2009). Note however, that using this criterion may inflate the number of frames for Hagar's corpus, which is considerably larger both than the other two samples as well as than previously used corpora (possible implications of this will be discussed later on). The same criteria were applied here to increase the comparability of this Hebrew-based study with results of relevant studies in other languages.

The current study differs in one way from earlier research in this area. Unlike previous studies, one-word frames were not included in the analysis, which took into account only two- and three-word frames, since concern here is with the kinds of grammatical relations conveyed by frames. The initial word of the utterance in Hebrew may convey little grammatical information, among other things, because of the fact that affixed functors are coded as independent words in this corpus, a procedure which would result in many one-word frames consisting of these functors (e.g., the definite article *ha-* 'the', prepositions *be-* 'in', *le-* 'to'). A preliminary analysis of all the utterance-initial words in the corpus (candidates for one-word frames) revealed that indeed, one-word frames carried little grammatical information. Most of the frequent utterance-initial words were either question words that were part of two-word frames (e.g., *ma-* 'what' appearing in *ma ze-*'what('s) that?'), or affixed functors, negation words, or connectors (e.g., *ve-* 'and') that do not provide children with grammatical information. Importantly, this procedure necessarily diminished the number of utterances counted as containing frequent frames.

As in previous studies, a single utterance could not be counted both as a two-word and a three-word frame. This study follows the procedure outlined in Stoll et al. (2009) for determining how to identify a frequent frame, illustrated here with an English example. After seeing the sequences 'where is mommy', 'where is the', 'where is your', and 'where is my' in one mother's speech, the sequence 'where is' is coded as a two-word frame (because it appears four times). If the sequence 'where is the' then appears three more times in the same mother's speech, the three-word frame 'where is the' will be added, and the two-word frame 'where is' will be removed. It would require another occurrence of 'where is', that is not followed by the word 'the', before the sequence 'where is' can be reclassified as a frequent two-word frame.

Table 1 illustrates the process applied to the Hebrew corpus in this study. If a particular mother's corpus included three instances of *ma at roca* 'what you-FM.SG want-FM.SG' = 'What do you want?' and one instance of *ma at mevina* 'what you-FM.SG understand-FM.SG' = 'What do you understand?' – the sequence *ma at* 'what you-FM.SG' would count as a two-word frame. However, if *ma at roca* 'what you-FM.SG want-FM.SG' was repeated again (to reach four repetitions), it would be introduced as a three-word frame, and that two-word frame *ma at-* would be removed until it was encountered three more times, without being followed by the word *roca*.

In order to see if different mothers use similar frames, the frequent frames were coded for generality of use: a frame was defined as a *core-frame* if it was used by at least two mothers (out of the three, Cameron-Faulkner et al. 2003). The frames were also coded manually for type to examine the kinds of relations children are exposed to in frequent-frames. Five such types were identified, based on those used by Cameron-Faulkner et al. 2003, with adaptations for Hebrew, as follows: (1) *Questions*: sequences containing question words like *ma at* 'what you-FM.SG' = 'What do

Table 1. Frequent Frame coding example

Utterances	Frame	Frame-length
1. *ma at roca?* **what you**-FM.SG **want**-FM.SG What do you want?		
2. *ma at xošévet?* **what you**-FM.SG **think**-FM.SG What do you think?		
3. *ma at roca?* What do you want?	*ma at* – what you-FM.SG	Two-word
4. *ma at roca?* What do you want?	*ma at roca* – what you-FM.SG	Three-word
5. *ma at roca?* What do you want?	want-FM.SG	

you?'. Yes/No questions were not included in this category since they would require manually examining the full utterance (including context) and not only the initial sequence; (2) *Subject-Predicate*: sequences containing a pronominal or lexical subject and a verb or an inflected verb in isolation like *ani avi lax* 'I bring-FUT.SG to-you-SG.FM' = 'I will bring you'. Because Hebrew is a (partially) pro-drop language where verbs can appear without a subject in certain paradigms (Berman 1990), a single verb can express subject-predicate relations; (3) *Embedded*: sequences containing the subordinating conjunction *še-* used in both complement and relative clauses like *at xošévet še* 'you-SG.FM think-PRS.SG.FM that' = 'you think that', *ra'íti yéled še-* 'see-PA.1SG. (a) boy-SG that' = 'I saw (a) boy that' respectively (Lustigman; Nir, this volume); (4) *Fragments*: sequences without a subject, a predicate, or a copula – such as prepositional phrases, noun phrases, discourse connectors (e.g., *aval ani* 'but I', *al ha-* 'on the'); and (5) *Routines*: sequences containing parts of songs, repetitions of words, and communicative routines (e.g., *láyla-tov* 'good night', *lo-lo-lo* 'No-no-no').

Results and discussion

The distribution of multiword information in Hebrew

The first analysis addressed the question of whether there are frequent frames in Hebrew CDS. A total of 25,716 utterance-initial sequences were counted in the speech directed to all three children. These utterances started with 13600 different trigrams and 7300 different bigrams. Of these, across the children, 257 were unique three-word frames (2%) and 610 unique two-word frames (8%) that appeared at the start of utterances – where a *frame*, as noted, was defined as a sequence that appeared at least four

times in the speech of one of the mothers. The term unique frame is used to indicate that if a frame appeared in the speech of more than one mother it was counted only once. The number of frequent frames differed between the mothers as a function of corpus size. Not surprisingly, the larger the corpus, the more likely it was for a sequence to be repeated at least four times. In terms of frequency, the frequent frames accounted for almost half of the utterance-initial sequences when looking across mothers, and between a third and a half of utterance-initial sequences when looking at each child separately (this calculation takes into account the different corpus sizes used for each mother). Put differently, a large proportion of the speech directed to children is accounted for by a relatively small set of repeated frames (see Table 2 for details). This analysis reveals that, as in other languages, the input to young Hebrew-speaking children contains many recurring multiword sequences despite its relatively rich morphological marking.

Table 2. Number of unique frames and their frequency for each mother

	Number of utterances	Number of unique frames	Frame token frequency	Proportion of utterances accounted for
Hagar	12,098	Three-word: 200 Two-word: 535	6,312	52%
Lior	3401	Three-word: 57 Two-word: 129	1353	40%
Smadar	1,600	Three-word: 19 Two-word: 63	500	32%
Total[2]	17,099	Three-word: 227 Two-word: 580	8163	48%

The analysis also reveals interesting differences between the mothers. In particular, there were more frequent-frames in the speech directed to Hagar, and those frames accounted for more of the utterances (larger coverage). One obvious explanation for this is the larger corpus size: Hagar's corpus was four times larger than Lior's and seven times larger than Smadar's. However, the picture did not change dramatically when the criterion for what counts as a frequent-frame was increased for this larger corpus. The number of frequent-frames in Hagar's mother's speech decreased when only frames that appeared *over ten times* across the corpus were included: there were now only 56 three-word frames, and 178 two-word frames. But their coverage remained high: They

2. Note that the total number of frames is not just a sum of the number of frames for each child, but rather the number of unique frames in the corpus (if a frame was found for two mothers, it was only counted once).

still accounted for 48% of the utterances directed to Hagar, suggesting that the speech directed to Hagar is characterized by a relatively small number of frequently repeated sequences. Another explanation for the high coverage in Hagar's corpus has to do with the particular interactional style of Hagar's mother, described in depth in Berman and Lustigman (2014). Hagar's mother is characterized as being very talkative, and using a somewhat intrusive interactional style which may result in more repetitions.

How consistent are these frames across mothers

To see how consistent frames are across different mothers, we looked at the amount of *core frames* used by each mother. A frame was defined as a core frame if it appeared in the speech of at least two of the three mothers. This yielded a total of 163 core frames in the corpus (including both two-word and three-word frames), suggesting that there is a high degree of consistency in the frames used with different children. Table 3 shows the number of core-frames for each child and the proportion of core-frames out of the frequent frames identified for each child.

Table 3. Number and proportion of core-frames for each mother

	Hagar	**Lior**	**Smadar**
Core Three-word Frames	38 (17%)	33 (58%)	11 (58%)
Core Two-word Frames	129 (24%)	119 (92%)	55 (87%)
All Core Frames	167 (22%)	157 (82%)	66(80%)

Many of the frames are repeated across children: over 80% of the frequent frames directed to Lior and Smadar are core frames, in contrast to only 20% of the frequent frames found in Hagar's corpus. This difference, again, probably reflects the larger size of the corpus: Because the corpus had four times more utterances than the other two, more idiosyncratic sequences were identified as frequent frames (including ones with the child's name in them, like *hagári ani roa* 'Hagari I see-PRS.SG.FM'). When we applied the more stringent definition of frequent-frame to Hagar's corpus (including only frames appearing over ten times per million), the proportion of core frames increased significantly, although remaining lower than that found for the other two children. Almost half of the frequent frames found in the corpus were now defined as core frames (112 out of 234, 48%), suggesting again that both corpus size and the idiosyncratic properties of Hagar's mother led to the observed differences between the children. The core-frame analysis shows that there is a quite a considerable overlap in the frequent frames used with different children, an important factor if frames are used to assist children in discovering grammatical relations and regularities.

How prevalent are frequent frames in Hebrew compared to other languages?
Before conducting a more in-depth analysis of the content of frequent frames in Hebrew, the number and coverage of frequent frames in Hebrew were compared to those found in other languages, where "coverage" refers to the number of utterances accounted for by the frequent-frames. Previous work has shown that the number and coverage of frames differs between languages, in line with their morphological richness and word order flexibility. Stoll et al.'s (2009) comparison of the number and coverage of frequent frames in English, German, and Russian showed, as expected, that English had on average more frames than German, which had more frames than Russian. A similar pattern was found for coverage: English frames accounted for more of the data than German frames, which in turn accounted for more of the data than Russian frames. In terms of length, English had more three-word frames than either of the other two languages, German had both three-word and two-word frames, while Russian had very few three-word frames.

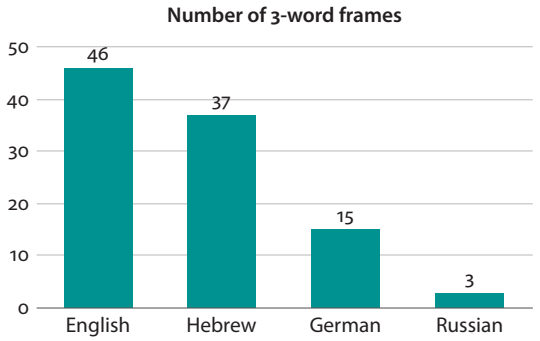

Figure 1. Average number of three-word frames in Hebrew, English, Russian and German, corrected for corpus size

Focus here is on the number of *three-word frames* in Hebrew compared with other languages, with the expectation that Hebrew would fall somewhere between German and English, since it is inflectionally richer than English but less so than German. For example, articles are not marked for gender, nor does it have such a complex case system – since in Hebrew case is inflected only on pronouns, not on lexical nouns. At the same time, Hebrew has more word-order variation than English but less than German or Russian. Figure 1 shows the number of three-word frames (on average) in Hebrew, after taking into account that the Hebrew corpus used here was larger than the ones used in comparable analyses of German and Russian. The combined corpus of three children used here had 17,000 child-directed utterances, while the study comparing English, German and Russian analyzed approximately 8,400 utterances for each language (1,400 for each mother). Several measures were taken to make the numbers

comparable. First, an average frame number was computed for each language by dividing the total number of three-word frames by the number of mothers (six mothers for German, English, and Russian, and three for Hebrew). The numbers for English, Russian, and German are taken from Stoll et al. (2009). Since this analysis did not extract one-word frames, but only two- and three- word frames, the comparison to other studies is not straightforward. Stoll et al. (2009) compare the total number of frames for English, German, and Russian respectively, without specifying exact numbers for two- and three- word frames (although they do examine the difference in frame length statistically). As a result, while the number of three-word frames in the different languages can be deduced from Figure 2, which gives the number for each mother, this is only an approximation. Second, the number of frames found in Hebrew was divided by two to reflect the fact that as corpus size increases so does the likelihood of a sequence repeating.

As predicted, Hebrew had more three-word frames than German and Russian, but fewer than English. Figure 2 shows the proportion of utterances accounted for by three-word frames in each language. For Hebrew, I calculated this by dividing the token frequency of all three-word frames (how many times each frequent frame appeared) by the total number of three-word sequences appearing in utterance initial position in child-directed speech. The numbers for English, Russian and German are estimated from Figure 3 in Stoll et al. (2009).

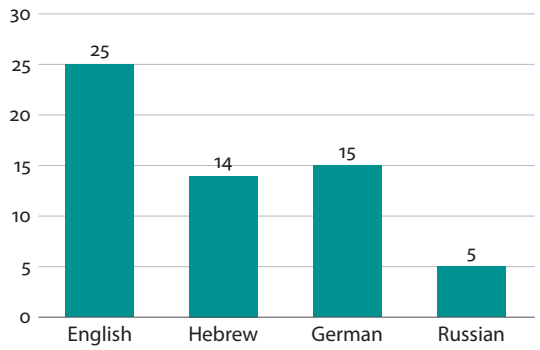

Figure 2. Proportion of utterances accounted for by three-word frames in Hebrew, English, Russian and German

As can be seen from Figure 2, the proportion of utterances accounted for by three-word frames in Hebrew is similar to that of German, lower than English, and higher than Russian. This pattern is compatible with the morphological and syntactic properties of Hebrew. The fact that Hebrew is inflectionally richer than English leads to it having fewer three-word frames, and to those frames accounting for less of the data.

However, having relatively fixed word order, with articles not marked for gender or number leads to more three-word frames than in German, and more coverage than was found for Russian.

In sum, the analysis of two- and three-word frames shows that (a) such frames are found in Hebrew, (b) their frequency and length reflect language-specific features, (c) they are repeated across speakers, and (d) they account for over half of the utterances children hear, highlighting the repetitive nature of child-directed speech and the prominence of multiword information that it contains.

What is the content of frequent frames in Hebrew?

This section provides a more in-depth analysis of the content of frequent-frames in Hebrew. As a first step, the different types of frames children hear were analyzed using the coding scheme outlined in the section on Methods. Frames were coded into five types intended to capture different syntactic constructions: Questions, Subject-Predicates, Fragments, Embedded, and Routines. Types were coded separately for two-word and three-word frames, as shown in Figures 3 and 4 respectively. Of the 257 unique three-word frames, the most frequent types were questions (42%), followed by subject-predicate frames like *at roca la'asot* 'you-SG.FM want-PRS.SG.FM to do-INF' 'you want to-do ~ make' (33%), and fragments (15%). Interestingly, 7% of the frames provided children with examples of embedded structures that are more syntactically complex like *ani xošévet še-*'I think-PRS.SG.FM that' (appearing 32 times in the corpus) or *at roca še-* 'you-SG.FM want-PRS.SG.FM that' (appearing at the prodigious rate of 110 times) in the sense of "do you want (someone) to …?". In fact, the proportion of Question frames may be even greater than that reported here since some of the Subject-predicate frames might actually be Yes-No questions.

The two-word frames showed a somewhat different distribution: Of the 580 unique two-word frames, the majority were fragments (52%) followed by subject-predicates and questions. The difference in distribution between the two- and three-word frames is not surprising: Three-word frames carry more grammatical information (and so can be classified into more types) simply by virtue of containing additional words/morphemes. The increase in fragments at the expense of questions and subject-predicates is a reflection of (a) the fact that many two-word frames consisted of two morphemes that might be better treated as one word (e.g., *ze ha-* 'this (is) the-'), and (b) the fact that many two-word fragments continue as questions or subject-predicates. For example, a sequence like *gam at* 'also you-SG.FM' = 'you, too' or *at lo* 'you-SG.FM not' = 'You don't' defined as a fragment even though it is most likely part of a question or subject-predicate sequence (e.g., *at lo boxa* 'you-SG.FM not cry-PRS.SG.FM' = "you aren't crying ~ you don't cry").

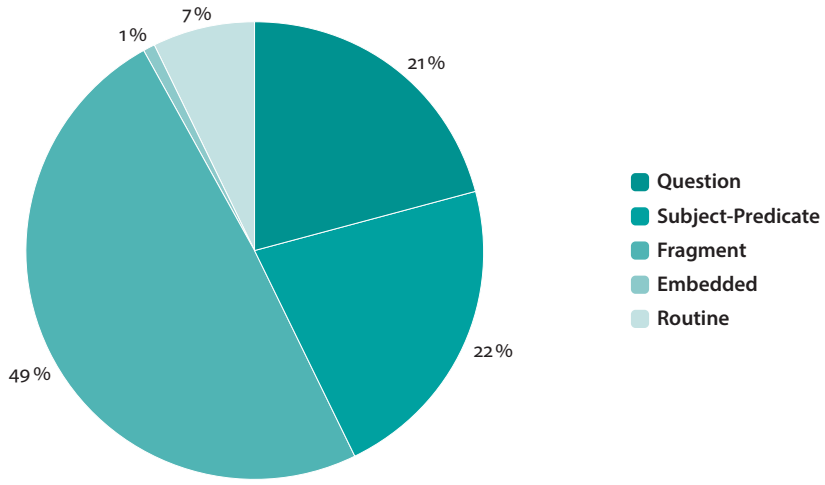

Figure 3. Distribution of two-word frames

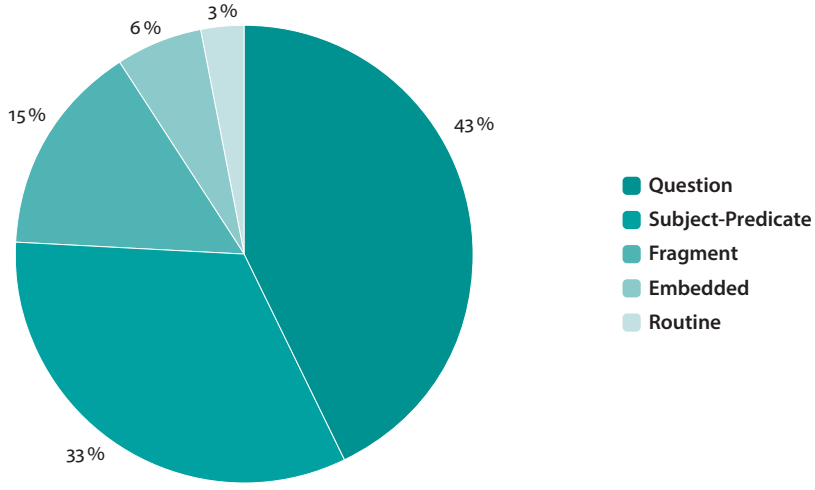

Figure 4. Distribution of three-word frames

Importantly, frequent frames in Hebrew present children with a range of syntactic structures and grammatical relations. Syntactically, the frames include simple clauses, questions, and embedded clauses. The simple clauses include a wide range of verbs (over 30 different verbs) and provide examples of different syntactic constructions, among them: various complex verb constructions (e.g., *at roca lecayer* 'you-SG.FM want-PRS.SG.FM to-draw-INF', *at roca lištot* 'you-SG.FM want-PRS.SG.FM to-drink-INF); including with aspectual verbs (e.g., *at gamart le'exol* 'you-SG.FM finish-2PST.SG.FM

to-eat-INF' = 'You've finished eating'); and existential constructions (e.g., *yeš li haca'a* 'be-PRS to-me suggestion-SG' = "I have a suggestion"). They also contain a number of frames illustrating negation in Hebrew (e.g., *at lo yoda'at* 'you-SG.FM not know-PRS.SG.FM' = 'You don't know", *ani lo marša* 'I not allow-PRS.SG.FM' = I don't allow"). The question-frames expose children to a full range of question words, starting with all of the following: question words (with numbers referring to frame types, not tokens): *ma* 'what' (55 frames), *mi* 'who' (8 frames), *láma* 'why' (6 frames), *ex* 'how' (4 frames), *éyze* 'which' (3 frames), *le'an* 'to-where' (3 frames), *éfo* 'where' (1 frame). By attending to recurring sequences, children can learn about the correct way of forming all these different question types. These frames – directed to children under 2;6 years of age – also include 17 different examples of embedded constructions like *at zoxéret še-* 'you-SG.FM remember-PRS.SG.FM that' indicating that from a young age, children are exposed to more complex syntactic constructions.

In terms of grammatical relations, frequent frames also provide children with examples of how to inflect different verbs: Over thirty different verbs participate in the frames, with many appearing in more than one inflection (past and present, singular and plural, masculine and feminine). For example, the root *r-a-y* 'see' appears in first, second and third person (*ani ro'a* 'I see-PRS.SG.FM', *at ro'a* 'you-SG.FM see-PRS.SG.FM, *hu ra'a* 'he see-PST.3SG.MS'); in singular and plural (*ro'im* 'see-PRS.PL.MS'); and in past, present, and future tenses (*tir'i* 'see-FUT.2SG.FM' = Imperative "look!"). Other verbs appear in similarly rich inflectional contexts. Subject-predicate frames provide children with information about the verbal inflectional paradigm, the nature of subject-verb agreement in Hebrew, as well as about the need to mark verbs and pronominal forms for gender, number and person.

Importantly, children are exposed to repeated occurrence of such relations. To demonstrate this, the 20 most frequent three-word core frames were considered, as listed in Table 4. The focus here is on core frames in order to increase the representativeness of the analysis, as representing sequences that appear frequently in speech directed to different children. In evaluating these data, the following should be noted. First, in terms of frequency of occurrence, all the sequences are highly frequent. The token frequency reported here is per 100,000 words (the size of the corpus). Each number needs to be multiplied by ten to make it comparable to the per million words measure that is generally used to assess frequency in psycholinguistic studies. In this literature, sequences appearing over ten times per million are considered to have a high frequency (e.g., Biber 2009). By this calculation, even the least frequent frame in the present sample turned out to have a frequency of over 100 occurrences per million words: In other words, children are very often exposed to the kinds of relations illustrated by these frequent frames.

Second, the frames illustrate a number of core syntactic and morpho-syntactic properties of Hebrew. They include a range of different verbs appearing in singular

Table 4. Twenty most frequent core frames

Frame	Token freq	Type
1. *ma at osa* what you-SG.FM do-PRS.SG.FM	151	Question
2. *ma at roca* what you-SG.FM want-PRS.SG.FM	121	Question
3. *at roca še-* you-SG.FM want-PRS.SG.FM that	110	Embedded
4. *at lo roca* you-SG.FM not want-PRS.SG.FM	61	Subject-predicate
5. *ani lo yodaat* I not know-PRS.SG.FM	44	Subject-predicate
6. *ve- ma od* and what else	42	Question
7. *at yodat ma* you-SG.FM know-PRS.SG.FM what	30	Question
8. *leʾan at holéxet* to-where you-SG.FM go -PRS.SG.FM	24	Question
9. *ma yeš lax* what there to-you-SG.FM = What have you got?	23	Question
10. *ex korʾim la-* how called-PRS.PL.MS to-the = What is the called?	22	Question
11. *at roca et* you-SG.FM want-PRS.SG.FM ACC	21	Subject-predicate
12. *ma at oméret* what you-SG.FM say-PRS.SG.FM	21	Question
13. *ani eten lax* I give-FUT.SG to-you-SG.FM	19	Subject-predicate
14. *ani lo mevina* I not understand-PRS.SG.FM	19	Subject-predicate
15. *ma kara la-* what happened-PST to-the	18	Question
16. *at xošévet še-* you-SG.FM think -PRS.SG.FM that	17	Embedded
17. *at roca laléxet* you-SG.FM want-PRS.SG.FM to-go	17	Imperative
18. *šel mi ha-* of who the- = Whose is the …	17 17	Fragment Imperative
19. *ani lo marša* I not allow-PRS.SG.FM	16	Question
20. *lama at lo* Why you-SG.FM not		

and plural, and in present and future tense. They illustrate subject-verb agreement for both gender and number. They include a range of syntactic constructions, including embedding, negations, and verb complements. The frames include many questions (as mentioned above, the number of question frames may be even higher since some of the other frames could be Yes-No questions), which are commonly found in child-directed speech (Cameron-Faulkner et al. 2003) and may play a role in facilitating interaction. These frames thus provide children with a rich source of information about the grammatical structure of Hebrew: they are both informative and repetitive, two features that are known to facilitate learning across domains (Ramscar, Yarlett, Dye, Denny & Thorpe 2010; Ramscar 2013). The frames also highlight the relationship between grammatical elements by exemplifying the kinds of agreement that a language marks (gender, number), a challenging and crucial aspect of language that children need to learn.

General discussion

The study aimed to investigate the nature and content of frequent frames in speech directed to young Hebrew-speaking children (under 2;6 years of age). Its goal was to examine the distribution, coverage, and content of such information to see what frames looked like in Hebrew compared to other languages – English, German, Russian (Stoll et al. 2009; Cameron-Faulkner et al. 2003) – and to ascertain whether they contained information that could be used to learn about the syntax and morphology of Hebrew. To this end, analysis was conducted of all the utterance-initial sequences from a large and dense corpus of speech directed to three different children as detailed in the Methods section (see, for example, Berman 1990; Berman & Lustigman 2014). The speech of each mother was examined separately to identify frequent two- and three- word frames (ones recurring at least four times across the corpus); each frame was then classified into one of five types (Subject-Predicate, Question, Embedded, Fragment and Routine); and "core" frames were then specified as those appearing in the speech of at least two of the three mothers. Following this, the number of frames, their coverage (the proportion of child-directed utterances they account for), and their content was analyzed. Looking across children, 227 unique three-word frames and 580 unique two-word frames were identified, which together accounted for over half of the child-directed utterances in the corpus. This number is similar to that found in other languages, disregarding one-word frames (which were not extracted in the present study).

This exploration shows that child-directed speech in Hebrew contains frequent frames, that such frames are relatively consistent across different caretakers, and that they account for a substantial proportion of child-directed utterances. The findings

highlight the repetitive nature of child-directed speech and the prominence of multiword information that it contains. Comparing the number and coverage of frames in Hebrew to other languages illustrates the effect of language-specific properties on the structure of child-directed speech. The degree to which utterance-initial sequences are repeated is dependent on word order variability – the more elements can appear in utterance-initial position, the fewer repeated sequences we will find. It is also affected by morphological richness: fewer frames can be expected in languages that have more morphological marking, meaning that sequences will differ depending on the properties of the participants. In line with these predictions, Hebrew – being an inflectionally rich language with relatively little word order variation – proves to have more two- and three-word frequent frames than German and Russian, but fewer than English.

Numerous frequent frames were identified in our corpus, indicating that Hebrew-speaking children are exposed to recurring multiword sequences. However, in order for this information to be useful for language learning, it needs to be both available and informative, appearing in sufficient quantities and illustrating relevant grammatical relations. Frequent-frames in Hebrew appear to fulfill both of these criteria. The frames appear frequently: the twenty most frequent frames appear over fifteen times in the corpus, and all of the frequent frames occur over four times. Given the size of the corpus, occurring four times in this Hebrew sample is equivalent to appearing forty times per million words, a frequency that is considered high in psycholinguistic investigations of multiword processing (Arnon & Snider 2010; Biber 2009). That is, children are exposed to frequent frames in quantities that are sufficient to support learning.

More importantly, frequent frames in Hebrew incorporate grammatically relevant information: they provide information about the grammatical relations children need to learn. The frames contain a range of different verbs and different syntactic constructions, including questions and embedded structures. They provide children with examples of subject-verb agreement, the verbal inflectional paradigm (in several tenses), and the obligatory marking of gender, person and number on verbs and pronouns. Frequent frames provide Hebrew-speaking children with rich and informative information about the grammatical regularities in their language. These findings are consistent with other studies of child-directed speech in Hebrew that highlight the structural information found in input to children (e.g., Berman 1990; Berman & Lustigman 2014; Ravid et al. 2009; Uziel-Karl 2006). More generally, the exploration of multiword information reveals that, as in other linguistic domains, children's input is richer and more informative than is often assumed (see Lieven 2014 for a recent discussion).

There are several ways in which the current study can (and should) be extended. First, the sample should include a mix of boys and girls. Looking only at speech directed to girls may have inflated the number of core frames, especially in a gender-marking language like Hebrew. Second, the study should be repeated with corpora that are more equal in size (Hagar's corpus was much larger than that of the other

two children), and that are more similar in size to other studies. It is possible that the number of frames (especially in Hagar's corpus) was inflated due to the bigger corpus size. Finally, and more importantly, the current study focused on what children hear. To convincingly argue that multiword information impacts learning, we need to show that such information influences child language. The relation between frequent frames in child and child-directed speech has been found in English (Cameron-Faulkner et al. 2003), but was not examined in the current study. However, there is growing evidence that children are sensitive to the distributional properties of multiword phrases. Young children (two- and three-year-olds) are faster and more accurate at repeating higher-frequency phrases (Bannard & Matthews 2008), while slightly older children (4;6) show better production of irregular plurals inside frequent frames (Arnon & Clark 2011). Children's early language illustrates their reliance on multiword information: many early utterances involve 'frozen' multiword chunks such as *give-me*, *get-it-out* (Lieven, Behrens, Speares, & Tomasello 2003; Lieven, Pine, & Baldwin 1997; Lieven et al. 2009), while later error patterns are impacted by multiword frequency (Dąbrowska & Lieven 2005; Rowland & Pine 2000). Me-for-I errors like *me-do*, for instance, are more frequent when children are exposed often to correct preverbal uses like *let me do it* (Kirjavainen, Theakston, & Lieven 2009). Given these findings, we expect to find a correlation between the frequent frames children hear, and the ones they produce.

In sum, this chapter focused on frequent-frames as defined in the usage-based tradition: repeated multiword sequences appearing at the start of utterances (Cameron-Faulkner et al. 2003). The study demonstrates that such frames are found in child-directed speech in Hebrew, that they account for a substantial proportion of child-directed utterances in the language, and that they provide children with relevant and informative grammatical information.

References

Abbot-Smith, K. & Tomasello, M. 2006. Exemplar-learning and schematization in a usage based account of syntactic acquisition. *The Linguistic Review* 23: 275–290. doi:10.1515/TLR.2006.011

Albert, A., MacWhinney, B., Nir, B. & Wintner, S. 2013. The Hebrew CHILDES corpus: Transcription and morphological analysis. *Language Resources and Evaluation* 47(4), 973–1005. doi:10.1007/s10579-012-9214-z

Armon-Lotem, S. & Berman, R. 2003. The emergence of grammar: Early verbs and beyond. *Journal of Child Language* 30(4): 845–877. doi:10.1017/S0305000903005750

Arnon, I. 2010. Starting Big: The Role of Multiword Phrases in Language Learning and Use. Ph.D. dissertation, Stanford University.

Arnon, I. & Snider, N. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62(1): 67–82. doi:10.1016/j.jml.2009.09.005

Arnon, I. & Clark, E.V. 2011. Why brush your teeth is better than teeth – Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development* 72: 107–129. doi:10.1080/15475441.2010.505489

Arnon, I. & Cohen Priva, U. 2013. More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech* 56(3): 349–371. doi:10.1177/0023830913484891

Arnon, I. & Cohen Priva, U. 2014. Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon* 9(3): 377–400. doi:10.1075/ml.9.3.01arn

Bannard, C. & Matthews, D. 2008. Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19(3): 241–8. doi:10.1111/j.1467-9280.2008.02075.x

Bannard, C., Lieven, E. & Tomasello, M. 2009. Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences of the United States of America* 106(41): 17284–9. doi:10.1073/pnas.0905638106

Berman, R. 1978. *Modern Hebrew Structure*. Tel Aviv: University Publication Project.

Berman, R. 1985. The acquisition of Hebrew. In *The Crosslinguistic Study of Language Acquisition* D.I. Slobin (ed.), 255–371. Hillsdale NJ: Lawrence Erlbaum Associates.

Berman, R. 1990. On acquiring an (S)VO language: Subjectless sentences in children's Hebrew. *Linguistics* 286: 1135–1166.

Berman, R. 2002. Crosslinguistic comparisons in later language development. In *The Diversity of Languages and Language Learning*, S. Strömqvist (ed.), 25–44. Lund: Center for Languages and Literature.

Berman, R. & Ravid, D. 2000. Research in acquisition of Israeli Hebrew and Palestinian Arabic. *Hebrew Studies* 41: 83–98. doi:10.1353/hbr.2000.0000

Berman, R. & Lustigman, L. 2014. Emergent clause-combining in adult-child interactional contexts. In *Language in Interaction. Studies in Honor of Eve V. Clark* [Trends in Language Acquisition Research 12], I. Arnon, M. Casillas, C. Kurumada, & B. Estigarribia (eds) 281–302. Amsterdam: John Benjamins. doi:10.1075/tilar.12.20ber

Biber, D. 2009. A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics* 14(3): 275–311. doi:10.1075/ijcl.14.3.08bib

Bod, R. 2009. From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science* 335: 752–93. doi:10.1111/j.1551-6709.2009.01031.x

Bracha Nir, B.M.S.W. 2010. A Morphologically-Analyzed CHILDES Corpus of Hebrew. *LREC*. ⟨http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.178.8392⟩

Bybee, J. 1998. The Emergent Lexicon. In *The 34th Chicago Linguistic Society* CLS 34: The panels, 421–435. Chicago IL: CLS.

Cameron-Faulkner, T., Lieven, E. & Tomasello, M. 2003. A construction based analysis of child directed speech. *Cognitive Science* 27(6): 843–873. doi:10.1016/j.cogsci.2003.06.001

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge MA: The MIT Press.

Culicover, P.W. & Jackendoff, R. 2005. *Simpler Syntax*. Oxford: OUP. doi:10.1093/acprof:oso/9780199271092.001.0001

Dąbrowska, E. & Lieven, E. 2005. Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16(3): 437–474. doi:10.1515/cogl.2005.16.3.437

Diessel, H. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 252: 108–127. doi:10.1016/j.newideapsych.2007.02.002

Elman, J. 2009. On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science* 33: 547–582. doi:10.1111/j.1551-6709.2009.01023.x

Givón, T. 1973. Complex NP's, resumptive pronouns, and word order in Hebrew. In *Papers from the Ninth Regional Meeting of the Chicago Linguistic Society*, C.T. Corum, T., C. Smith-Stark & A. Wieser (eds),135–146. Chicago IL: CLS.

Givón, T. 1976. On the VS order in Israeli Hebrew: Pragmatics and typological change. In *Studies in Modern Hebrew Syntax and Semantics*, P. Cole (ed.), 153–181. Amsterdam: North-Holland.

Goldberg, A. 2006. *Constructions at Work*. Oxford: OUP.

Huttenlocher, J., Vasilyeva, M., Cymerman, E. & Levine, S. 2002. Language input and child syntax. *Cognitive Psychology* 45(3): 337–374. doi:10.1016/S0010-0285(02)00500-5

Kirjavainen, M., Theakston, A. & Lieven, E. 2009. Can input explain children's me-for-I errors? *Journal of Child Language* 36(5): 1091–114. doi:10.1017/S0305000909009350

Langacker, R.W. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford CA: Stanford University Press.

Lieven, E.V.M., Pine, J.M. & Baldwin, G. 1997. Lexically-based learning and early grammatical development. *Journal of Child Language* 24(1): 187–219. doi:10.1017/S0305000996002930

Lieven, E., Behrens, H., Speares, J. & Tomasello, M. 2003. Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30(2): 333–370. doi:10.1017/S0305000903005592

Lieven, E., & Tomasello, M. 2008. Children's first language acquisition from a usage-based perspective. In, *Handbook of Cognitive Linguistics and Second Language Acquisition*, P. Robinson & N.C. Ellis (eds), 168–196. London: Routledge.

Lieven, E., Salomo, D. & Tomasello M. 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics* 20(3): 481–508. doi:10.1515/COGL.2009.022

Lieven, E. 2010. Input and first language acquisition: Evaluating the role of frequency. *Lingua* 120(11): 2546–2556. doi:10.1016/j.lingua.2010.06.005

Lieven, E. 2014. First language development: A usage-based perspective on past and current research. *Journal of Child Language* 41(Suppl. 1): 48–63. doi:10.1017/S0305000914000282

MacWhinney, B. 2000. *The CHILDES Project: Tools For Analyzing Talk*. Mahwah NJ: Lawrence Erlbaum Associates.

McClelland, J.L. 2010. Emergence in cognitive science. *Topics in Cognitive Science* 2(4): 751–770. ⟨http://philpapers.org/rec/MCCEIC⟩ doi:10.1111/j.1756-8765.2010.01116.x

Mintz, T.H. 2003. Frequent frames as a cue for grammatical categories in child directed speech. *Cognition* 90(1): 91–117. doi:10.1016/S0010-0277(03)00140-9

Naigles, L.R. & Hoff-Ginsburg, E. 1998. Why are some verbs learned before other verbs? Effects of input frequency and structure on children's early verb use. *Journal of Child Language* 25(1): 95–120. doi:10.1017/S0305000997003358

Peters, A.M. 1977. Language learning strategies: Does the whole equal the sum of the parts? *Language* 53(3): 560–573. doi:10.2307/413177

Peters, A.M. 1983. *The Units of Language Acquisition*. Cambridge: CUP.

Pinker, S. 1999. *Words and Rules: The Ingredients of Language*. New York NY: HarperCollins.

Ramscar, M., Yarlett, D., Dye, M., Denny, K. & Thorpe, K. 2010. The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science* 34(6): 909–957. doi:10.1111/j.1551-6709.2009.01092.x

Ramscar, M. (2013). Suffixing, prefixing, and the functional order of regularities in meaningful strings. *Psihologija* 46(4): 377–396. ⟨http://www.doiserbia.nb.rs/Article.aspx?ID=0048-57051304377R&AspxAutoDetectCookieSupport=1#.VPnUArOUe5I⟩ doi:10.2298/PSI1304377R

Ravid, D., Dressler, W.U., Nir, B., Korecky-Kröll, K., Souman, A., Rehfeldt, K., Laaha, S., Bertl, J., Basbøll, H. & Gillis, S. 2009. Core morphology in child directed speech: Crosslinguistic corpus analyses of noun plurals. In *Corpora in Language Acquisition Research: History, methods, perspectives* [Trends in Language Acquisition Research 6], H. Behrens (ed.), 149–162. Amsterdam: John Benjamins. doi:10.1075/tilar.6.05rav

Reali, F. & Christiansen, M.H. 2007. Word chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology (2006)* 60(2): 161–70. doi:10.1080/17470210600971469

Rowland, C.F. & Pine, J.M. 2000. Subject–auxiliary inversion errors and wh-question acquisition: "What children do know?" *Journal of Child Language* 27(01): 157–181. doi:10.1017/S0305000999004055

Seidl, A. & Johnson, E.K. 2006. Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental Science* 9(6): 565–573. doi:10.1111/j.1467-7687.2006.00534.x

Stoll, S., Abbot-Smith, K. & Lieven, E. 2009. Lexically restricted utterances in Russian, German, and English child-directed speech. *Cognitive Science* 33(1): 75–103. doi:10.1111/j.1551-6709.2008.01004.x

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. doi:10.1017/S0305000900004608

Tomasello, M. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge MA: Harvard University Press.

Tremblay, A., Derwing, B., Libben, G. & Westbury, C. 2011. Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning* 61(2): 569–613. doi:10.1111/j.1467-9922.2010.00622.x

Uziel-Karl, S. 2001. *A Multidimensional Perspective on the Acquisition of Verb Argument Structure*. Tel-Aviv: Tel-Aviv University.

Uziel-Karl, S. 2006. Acquisition of verb structure from a developmental perspective: Evidence from child Hebrew. In *The Acquisition of Verbs and their Grammar: The Effect of Particular Languages*, N. Gagarina & I. Gülzow (eds), 33: 15–44. Dordrecht: Springer. doi:10.1007/978-1-4020-4335-2