

ARTICLE

# The Starting Big approach to language learning

Inbal ARNON 

Psychology Department, Hebrew University, Jerusalem

Address for correspondence: Inbal Arnon, Psychology Department, Hebrew University, Jerusalem, ISRAEL.

Email: [Inbal.arnon@mail.huji.ac.il](mailto:Inbal.arnon@mail.huji.ac.il)

(Received 17 February 2020; revised 9 December 2020; accepted 4 May 2021;  
first published online 5 July 2021)

## Abstract

The study of language acquisition has a long and contentious history: researchers disagree on what drives this process, the relevant data, and the interesting questions. Here, I outline the Starting Big approach to language learning, which emphasizes the role of multiword units in language, and of coarse-to-fine processes in learning. I outline core predictions and supporting evidence. In short, the approach argues that multiword units are integral building blocks in language; that such units can facilitate mastery of semantically opaque relations between words; and that adults rely on them less than children, which can explain (some of) their difficulty in learning a second language. The Starting Big approach is a theory of how children learn language, how language is represented, and how to explain differences between first and second language learning. I discuss the learning and processing models at the heart of the approach and their cross-linguistic implications.

**Keywords:** language acquisition; usage-based models; multiword units

## 1. Introduction

The study of language acquisition has a long and contentious history: researchers disagree on what drives this process, the relevant data, and the interesting questions. Some of these disagreements stem from different assumptions about what it means to know language. While often unstated, the conceptualization of linguistic knowledge influences how we do language acquisition research: our definition of the adult state (what is known by the proficient speaker) critically shapes what we think children need to learn, and how much we think they differ from adults in their knowledge. For example, if we think that knowing language means knowing abstract syntactic rules whose use is not impacted by lexical frequency, we will search for such abstraction in children's early language. Alternatively, if we think that syntactic knowledge is probabilistic and tied to usage-patterns, we will search for frequency effects on children's early syntactic constructions. Our evaluation of child language, and its similarity to adult language, will be very different depending on how we characterise what it means to know language. To provide a comprehensive and testable theory of language acquisition we need to make our assumptions about the

© The Author(s), 2021. Published by Cambridge University Press

adult state explicit: the strength of our theory should be evaluated on the basis of the claims we make about how children learn language AND about how linguistic knowledge is represented and processed by the proficient adult speaker. Making these assumptions explicit can help us identify additional differences between theories of language acquisition – for instance, in their models of processing and representation – and better evaluate them by drawing on converging evidence from both children and adults.

In this spirit, before I present the Starting Big approach to language learning, I will outline some of the theoretical assumptions that lie at the core of this approach. The first, and most fundamental tenet, is that language acquisition and language processing are inherently linked. Language learning is viewed as the continuous process of learning to predict and learning to engage in interaction. This process does not stop when the child reaches a certain age – we are constantly updating our linguistic knowledge and expectations based on prior and recent experiences. The second tenet is that learning language has both domain-specific and domain-general features: while the learned content is unique, at least some of the learning mechanisms used to acquire it are domain-general, and can be used to learn other kinds of regularities as well. A good example comes from the domain of statistical learning: the ability to detect recurring patterns in the input is domain-general and implicated in learning both linguistic and non-linguistic stimuli (see Saffran & Kirkham, 2018) for a recent review). This domain-general perspective makes findings from other developmental fields highly relevant for the study of language acquisition: changes in memory, attention, and even motor skills are expected to interact with and influence children’s linguistic development. The third tenet is that language learning and processing are fundamentally impacted by experience: both our linguistic representations, and the mechanism we use to acquire them, are influenced by what we already know. Prior knowledge will impact how we learn (e.g., by diverting attention to word units or multiword ones, as I will discuss below), and how we structure our current and future representations.

Both language acquisition and language processing are seen as guided and shaped by similar predictive processes, some of which may be unique to language, while others are domain-general. While the initial process of acquiring linguistic knowledge has unique characteristics not found in real time processing (e.g., the need to discover linguistic units, to learn meanings, to learn grammatical constraints), the prediction is that there will be continuity in the factors impacting learning and processing. This prediction reflects a philosophical preference for simpler theories (Occam’s razor): an account that explains child and adult linguistic behaviour in the same way is more parsimonious than an account that provides two different explanations. A similar stance is taken towards innateness: postulating innate linguistic knowledge should be a last resort, acceptable only if the data cannot be explained otherwise. Here also, the preference is to explain language learning patterns using mechanisms and models that are independently motivated by adult psycholinguistic findings, or by developmental data from other domains (social cognition, memory, vision, etc.).

This assumption of continuity makes adult psycholinguistic findings highly relevant for the study of child language acquisition: such findings provide a way of generating predictions and providing explanations for developmental patterns. The utility of this approach can be exemplified by looking at the acquisition of relative clauses. It has long been noted that children struggle with comprehending object relative clauses (e.g., I saw the man that the boy chased), and find them harder than subject relative

clauses (e.g., I saw the man that chased the boy). In lab-based settings, even five-year-olds show poor comprehension of these constructions (e.g., Kidd & Bavin, 2002). This difficulty was claimed to reflect children's syntactic immaturity, and specifically their underdeveloped ability to process sentences that involve syntactic movement (which object relatives are thought to contain, (Friedmann & Novogrodsky, 2004)). However, this syntactic immaturity account is undermined when psycholinguistic and corpus-based findings are taken into consideration: children produce object relative clauses spontaneously from early on (around age two, (Diessel & Tomasello, 2001)), indicating they have some mastery of their syntax. Moreover, adults also find object relative clauses harder, despite the fact that their syntactic abilities are fully developed. In the adult literature, this increased difficulty is attributed to the greater processing cost associated with object relative clauses, because of the longer distance between the head and the verb, and the need to process an additional noun-phrase before resolving the dependency (e.g., Gibson, 1998).

Applying the same explanation to children would predict that their difficulty also reflects increased processing costs and not just immature syntax: children's comprehension should be facilitated when presented with object relative clauses that are easier to process. Indeed, children show improved comprehension when presented with object relative clauses that have pronominal subjects (e.g., The man that I chased) or inanimate heads (e.g., The ball that I chased, (Arnon, 2010; Kidd, Brandt, Lieven & Tomasello, 2007)). These configurations are also ones that children actually hear: object relative clauses with two animate lexical NPs (like those often presented in experimental settings, *The boy that the man chased*) are vanishingly rare in both child-directed and adult-to-adult speech (Real & Christiansen, 2007a). That is, processing cost and input frequency can explain child and adult difficulty without having to postulate additional constraints (like immature syntax). This discussion of relative clause acquisition illustrates the theoretical and methodological tenets that underlie my approach to language learning. Because there is continuity in the factors impacting learning and processing, it is both necessary and advantageous to use converging evidence from child language and adult processing when formulating theories of language acquisition. Now that I've made these theoretical assumptions explicit (and therefore subject to scrutiny), I turn to the description of the Starting Big approach itself.

## 2. The Starting Big approach to language learning

The classic textbook description of language acquisition often emphasizes children's transition from smaller linguistic units to larger and more complex ones, going from the production of syllables (babbling), to single words, followed by multiword combinations, and finally complex sentences. The Starting Big approach highlights the importance of another, complementary, process: the move from larger and more holistic units to smaller and fully analysed ones. In particular, the approach highlights the role of multiword units in language learning and use. Multiword units are defined as sequences larger than a single lexical word, that have a representation in the mental lexicon (the term multi-morphemic may be more appropriate cross-linguistically, see discussion in section 4). Having a lexical representation means that speakers store information about the whole multiword unit alongside information about its parts. For example, speakers will represent the frequency of the entire multiword sequence *don't have to worry*, in addition to the frequency of the individual words and sub-strings that

it contains (*don't, have, to, worry, don't have, have to, to worry, don't have to, have to worry*). This does not mean that all multiword sequences result in a multiword representation: there has to be a reason for forming a larger representation (frequency, meaning, prosody, e.g., Jolsvai, McCauley & Christiansen, 2020). It also does not mean that multiword units are stored holistically without access to their component parts (Arnon & Cohen Priva, 2014; Baayen, Hendricks & Ramscar, 2013; Bolinger, 1968; Siyanova-Chanturia & Martinez, 2014).

Acknowledging the presence and prevalence of multiword units in language has implications for understanding first language acquisition, second language learning, and the differences between them. In a nutshell, I propose that children draw on multiword units in the learning process and use them to learn grammatical relations between words. Such units become part of the developing lexicon via under-segmentation or chunking, and remain as an integral part of the native adult lexicon, impacting language processing and representation (see section 3.1). Learning from them can enhance mastery of semantically opaque grammatical relations that hold between words (see section 3.2). Importantly, under the Starting Big approach, adult learning a second language are expected to rely less on multiword units because of their existing knowledge of words, and the fact that they are usually literate, both of which increase attention to word-level units (see section 3.3). Adults' smaller reliance on multiword units is predicted to hinder their mastery of semantically opaque grammatical relations between words (see section 3.2). The Starting Big approach is a theory of how children learn language, how language is represented, and how we can explain differences between first and second language learning. It aims to provide a unified, experience-based explanation for findings from child language, adult psycholinguistics, and second language learning. In the following sections I outline the core predictions of the theory, alongside supporting evidence and open challenges, and then turn to its' ability to explain core phenomena in the acquisition of syntax and morphology across languages.

## 2.1 Multiword units as building blocks in first language learning

Children's early lexical inventories are predicted to contain a mix of single words and short multiword sequences (e.g., *what's-that*). This is seen as a natural consequence of their developing perceptual abilities and the nature of their input. Because infants do not initially know where word boundaries are (or even that words exist as units of representation), and because they can perceive larger unit boundaries before smaller ones (e.g., utterance boundary before word boundary, (Soderstrom, 2003)), their initial linguistic units are extracted on the basis of prosodic, not lexical boundaries. Given the nature of child-directed speech, these will include both single words and short multiword sequences (Cameron-Faulkner, Lieven & Tomasello, 2003). As they accumulate linguistic experience, children will segment these initially under-segmented chunks. We can distinguish between two types of multiword units (Arnon & Christiansen, 2017): (1) multiword units formed via under-segmentation (described above), which are only formed in the early stages of language acquisition, and (2) multiword units created via chunking, where words that co-occur together often (or have other binding properties like meaning or prosody) give rise to a multiword representation (which exists alongside the representation of the individual words). The creation of multiword units via chunking is expected to continue throughout the life span. That is, multiword units are seen as an integral part of what native speakers know about their language (e.g., Arnon & Snider, 2010).

This view of linguistic representation is very different from that of dual-system models. By hypothesizing that the lexicon contains both words and multiword units, all processed and represented using the same mechanisms, this theory is distinct from dual-system models, of which words-and-rules (Pinker, 1999) is the most famous instantiation. In dual-system models, knowing language means knowing the atomic elements (words, morphemes), and the rules or constraints used to combine them. There is a clear (and qualitative) separation between forms that are stored in the lexicon (morphemes, words, irregular forms, idiomatic expressions) and forms generated by the grammar (e.g., regular forms, compositional phrases): the two are said to be processed differently, represented differently, and even to have distinct neural realizations (Ullman, 2004). Single-system/emergentist models, in contrast, assume that all linguistic experience is processed by a single cognitive mechanism. While there are many different implementations of such models (construction grammar – Goldberg, 2006); connectionist – Rumelhart & McClelland, 1986; exemplar-based – Bod, 2006), none of them advocate a clear separation between lexicon and grammar, or between stored and computed forms. Instead, the lexicon is thought to contain units of varying sizes and levels of abstraction. The representation and learning of multiword sequences is naturally accommodated in such models: as long as multiword sequences are frequently experienced, they will be represented in memory, and will impact both production and comprehension. The degree to which they will be perceived as one unit will depend on their frequency (overall, and relative to the frequency of their parts); their content (how meaningful they are as a unit, (Jolsvai, McCauley & Christiansen, 2013, 2020); and their function (do they convey a social or discursive function). Moreover, because no qualitative difference is predicted in how we process different aspects of linguistic experience, the processing of multiword sequences should be subject to the same constraints that impact the processing of single words.

Indeed, there is growing evidence that children and adults represent and draw on multiword units in learning and processing, and treat them as they treat lexical words. Children and adults are sensitive to multiword frequency: adults are faster to produce and comprehend higher frequency multiword sequences compared to lower frequency ones (Arnon & Snider, 2010; Reali & Christiansen, 2007b), and show better memory for them (Tremblay, Derwing, Libben & Westbury, 2011). These effects hold when controlling for all part frequencies and are also found in spontaneous speech: for example, the phonetic duration of three-word sequences in conversation is shorter when they are more frequent (Arnon & Cohen Priva, 2013, 2014). Children show similar patterns: two-year-olds are faster and more accurate in repeating higher frequency four-word phrases (Bannard & Matthews, 2008), and slightly older children show better production of irregular plurals when produced as part of a more frequent phrase (e.g., Brush your --- teeth vs. So many --- teeth, (Arnon & Clark, 2011). Such findings suggest that native speakers (children and adults) represent multiword units in addition to words.

However, they do not show that such units are used as building blocks during first language acquisition. It is not easy to provide evidence for the role of larger units in the early stages of first language acquisition. By the time they start to talk, children have already done a lot of analysis and segmentation: their units of production are not necessarily the same as their units of perception (Clark & Hecht, 1983). We can see this clearly in the domain of word learning: while children only start producing their first words around 12-months, by six-months they are capable of segmenting words

from fluent speech (Jusczyk, 1999), and have a few rudimentary object-label associations (Bergelson & Swingley, 2012). There is also divergence between children's phonetic representations for production and perception that is often referred to as the *fis*-phenomenon (Dodd, 1975). This describes the situation where children prefer the adult form in perception, even when their own productions are not adult-like. For instance, a child producing [dup] for jump will nevertheless prefer to listen to the adult-like *jump*, and will protest if the adult uses the child form (Clark, 1978). Our understanding of lexical acquisition would be inaccurate if we based it on production data alone. Similarly, the fact that children's early productions are mostly single-word utterances, does not mean that they do not extract and represent larger units as well: their absence in production could reflect articulatory difficulty rather than representational absence, as well as researchers' bias for finding words and not larger sequences.

In a seminal study, Ann Peters (Peters, 1977), closely examined the early speech of one child, and reported the presence of what she called Gestalt utterances. These utterances had the intonational contour of an adult phrase but did not contain identifiable 'word' units. At age 1;2, for example, the child produced /obe-da-do/ while banging on the bathroom door (to mean *open the door*). At this point, the child had only 10 recognizable words in his productive vocabulary, not including either *door* nor *open*. At 1;7 the same child made frequent use of /siliini?/ to mean *silly, isn't it?*. The word *silly* did not appear in other utterances (or on its own) until much later. Instead of producing the utterance by combining known words, the child seemed to be producing an unanalyzed, or under-analyzed, multi-word chunk. Building on this case-study and several others, Peters proposed that children's early communicative attempts consist of adult-like prosodic contours that are only later segmented into recognizable words, and that many of children's early units may initially contain more than one lexical word (Peters, 1983). She also noted that children's early multiword utterances may go undetected by researchers (and parents) who look for individual words. Peters insightful account was based on the speech of very few children, and unfortunately, few studies have followed up on her work to provide a more quantitative investigation of multiword units in children's early inventory.

So how can we nevertheless identify larger units in early child language? This challenge can be approached using computational and experimental tools. Children's early productions are better accommodated by a computational model that extracts both words and multiword units (McCauley & Christiansen, 2017, 2019). In another study, (Borensztajn, Zuidema & Bod, 2009) use data-oriented parsing, to identify the most likely primitive units in early child speech. They conclude that many of them are multiword fragments. While this does not prove that children draw on multiword units, it shows that the extraction of larger units is supported by input patterns, and is useful for explaining early child productions. More direct evidence for the presence of larger units or chunks in the early lexicon is provided by looking at adult processing and infant speech perception. If multiword units are part of the developing lexicon, then they may leave traces in the adult lexicon, as has been found for single words. Lexical Age-of-Acquisition effects have been found in numerous studies: words that were acquired earlier show processing advantages in adult speakers (see Juhász, 2005) for a review). This advantage is interpreted to mean that early-acquired words leave traces in the adult lexicon. Extending this logic, we find Age-of-Acquisition effects for three-word sequences: like words, sequences that

were acquired earlier (based on corpus-data and speaker ratings) are processed faster by adult speakers compared to later-acquired ones (I. Arnon, McCauley & Christiansen, 2017). To give an example, the trigram ‘*a good girl*’ is very frequent in child-directed speech and rated as early-acquired, while the almost identical sequence ‘*a good dad*’ is very infrequent in the same child-directed corpus and is rated as later acquired. Adults show faster response times to the early-acquired sequence compared to the later-acquired one (after controlling for lexical AoA, plausibility, and all part and whole frequencies in adult speech). That is, like words, early-acquired sequences have a privileged status in the adult lexicon, supporting their role as early building blocks.

A second processing parallel between words and larger units is found when we look at infant speech perception. Infants are sensitive to word frequency (Bergelson & Aslin, 2017). If they extract multiword units and process them similarly, infants should also be sensitive to the frequency of larger sequences. To test this, we used an infant-controlled sequential looking procedure to ask if 11-month-olds are sensitive to multiword frequency (Skarabela, Ota, O’Connor & Arnon, 2021). We compared looking times to trigram pairs that differed in only one word, had similar lexical frequency, but different trigram frequency (calculated from a corpus of infant-directed speech). For example, the trigram *clap.your.hands* is very frequent in infant-directed speech while the almost identical *take.your.hands* is not. Infants looked significantly longer when hearing the frequent trigrams compared to the infrequent ones, suggesting they are sensitive to frequency differences in multiword combinations. This is the first demonstration, to our knowledge, of infants’ sensitivity to the distributional properties of larger sequences at a stage when they are barely producing single words.

### 3. The impact of multiword units on learning: some core predictions

The above-mentioned findings suggest that children use multiword units as building blocks in the process of language acquisition, and that adult native speakers continue to represent larger units. But how precisely do multiword units support learning? The idea that larger patterns help children learn about language structure is not a new one: it is a core prediction of usage-based models of language acquisition (Tomasello, 2003). In such models, grammatical knowledge is acquired by abstracting over stored exemplars (Abbot-Smith & Tomasello, 2006; Bod, 2009). Children start off with signs – concrete, lexically-realized linguistic patterns – which can be single words (e.g., *mommy*) or multiword utterances (e.g., *what is that?*). These units are abstracted over to create schemas – partially realized frames with more abstract slots (e.g., *what is X?*), which eventually give rise to more abstract knowledge (Lieven, Behrens, Speares & Tomasello, 2003; Lieven & Tomasello, 2008; Lieven, Pine & Baldwin, 1997). Multiword units play an important role in such models: they provide children with building blocks from which grammatical relations between words can be learned.

Indeed, children’s input contains many recurring multiword sequences, often referred to as frequent frames. Language acquisition studies highlights at least two types of frequent frames: the usage-based literature looks at CONTINUOUS sequences, often appearing sentence-initially (such as *Are you ---*, *That is ---*, Cameron-Faulkner et al., 2003). The literature on grammatical category acquisition uses the term to refer to NON-CONTINUOUS sequences: “two jointly occurring words with one word intervening” (Mintz, 2003) like *the – is*, or *you – it*. Both types of frames are thought to facilitate language learning. Continuous frequent frames are prevalent in

child-directed speech, across languages, though their characteristics are impacted by language specific features. Russian, for instance, has fewer frequent frames compared to German and English, which is expected given Russian's more flexible word order and richer inflectional system (Stoll, Abbot-Smith & Lieven, 2009). Languages with more inflectional marking (e.g., for gender and number) will necessarily have fewer repeated sequences compared to languages with less morphological marking, since the same utterance will have different forms, depending on the gender, number, and person of the participants. Reflecting this, Hebrew, an inflectionally rich language with relatively fixed word order had more frequent frames than found in Russian and German (which have more flexible word order), but fewer than English (which has less morphological marking, Arnon, 2016). Frequent frames could help children learn the words that appear in them, expose them to a diverse range of constructions (questions, imperatives, complement clauses), and provide them with multiword sequences to be used in their own productions. For example, the distribution of verb forms appearing after frequent frames in Hebrew is representative of child-directed speech as a whole, and can be used to extract accurate root-based information (Johnson & Arnon, 2019).

Non-continuous frequent frames provide children with a different kind of information. Such frames can serve as a cue for word categorization: words that appear in the middle slot of frequently occurring frames tend to be of the same grammatical category (e.g., noun, verb). By clustering these words together, based on their appearance in a discontinuous frequent frame, children can start building grammatical categories. Indeed, applying this clustering method to English child-directed speech, results in the formation of mostly accurate grammatical categories (Mintz, 2003). Subsequent studies investigated the utility of such frames cross-linguistically, with mixed results: non-adjacent frames provided a good cue for category formation in some languages (Spanish – Weisleder & Waxman, 2010; French – Chemla, Mintz, Bernal & Christophe, 2009), but not in others (Turkish – Wang, Höhle, Ketrez & Küntay, 2011; Dutch – Erkelens, 2009). The degree to which they did depends on the morphological richness of the language, and whether words or morphemes are used as the unit of analysis (word1 --- word2 vs. morpheme1 – morpheme2). A recent analysis of seven typologically different languages – Chintang, Inuktitut, Japanese, Russian, Sesotho, Turkish, and Yucatan – shows that non-adjacent frames do indeed provide a reliable cue to category formation across languages, once language properties are taken into account (Moran et al., 2018). The authors evaluate the goodness of frames as a cue for 12 grammatical categories: frequent frames at the word level were not a good predictor cross-linguistically, but frames at the morpheme level were. Taken together, the literature on continuous and non-continuous frames illustrates the prevalence of recurring multiword sequences across languages, and points to their facilitative role in learning grammatical regularities.

The Starting Big approach draws on this evidence, and on usage-based models more generally, to make several novel predictions about the impact of larger units on language learning, and their differential role in first and second language learners. The first prediction is that MULTIWORD INFORMATION IMPACTS LEXICAL AND MORPHOLOGICAL ACQUISITION. The second is that MULTIWORD UNITS FACILITATE MASTERY OF SEMANTICALLY OPAQUE GRAMMATICAL RELATIONS, the third prediction is that PRIOR KNOWLEDGE AND EXPERIENCE IMPACT RELIANCE ON MULTIWORD UNITS. Together, these lead to the prediction that: ADULTS, BECAUSE OF THEIR PRIOR KNOWLEDGE, RELY LESS ON MULTIWORD UNITS IN



LEARNING, WHICH CAN EXPLAIN (SOME OF) THE DIFFICULTY THEY HAVE IN LEARNING A SECOND LANGUAGE. I will present the predictions below, alongside the existing evidence supporting them, and outline additional ways they can (and should) be tested.

### 3.1 Multiword information impacts lexical and morphological acquisition

Numerous studies illustrate the effect of input frequency on language learning: the sounds, words, and constructions that appear more often in child-directed speech, tend to be acquired earlier (Ambridge, Kidd, Rowland & Theakston, 2015). Moreover, the amount (and quality) of input children hear has cascading effects on their linguistic abilities (though see Cristia, Dupoux, Gurven & Stieglitz, 2019 for a critique of whether these findings generalise beyond WEIRD societies). They impact vocabulary size (Hart & Risley, 1995; Mahr & Edwards, 2018), processing speed (Fernald, Marchman & Weisleder, 2013), and subsequent literacy acquisition (Noble, Farah & McCandliss, 2006). Such frequency effects provide support for usage-based models, where – in contrast with Universal Grammar (Chomsky, 1965) or other accounts that emphasize innate linguistic knowledge (Hyams, 1988) – input frequency is seen as one of the main forces driving language acquisition.

Under the Starting Big approach, learners are predicted to be sensitive to both word and multiword frequency, and to draw on multiword information in discovering and learning about smaller units (words, morphemes). Adopting a single-system view of linguistic representation – as the Starting Big approach does – means that words are represented alongside larger patterns and are connected to the larger patterns they appear in, with consequences for learning and processing. For example, it is easier to access words when they appear as part of a more frequent multiword sequence (e.g., Arnon & Snider, 2010). This model has developmental consequences: the prediction is that children's knowledge of smaller units (morphemes, words) is not independent from the larger patterns they appear in. Children's ability to produce or comprehend words and morphemes will be modulated by the larger linguistic context they appear in. Linguistic knowledge is not seen as an all-or-nothing state, but as a gradient one where the ability to use the correct form depends (among other things) on the immediate linguistic context. That is, children's ability to use certain morphemes and words will reflect and be impacted by multiword information.

This can be illustrated using one of the most studied domains in language acquisition: the acquisition of irregular plurals. Children have difficulty acquiring irregular forms, and often produce both correct and over-regularized forms (using *mouses* for *mice*). The source of these errors has been heavily debated: do children's over-regularization errors reflect the application of an abstract rule (e.g., Marcus et al., 1992), or the higher activation of the more frequent regular plural marking (e.g., Rumelhart & McClelland, 1986)? Much work has examined the distributional, semantic and phonological factors that impact the frequency of such errors (e.g., Maslen, Theakston, Lieven & Tomasello, 2004; Matthews & Theakston, 2006). All these studies, however, focus on word-level properties, showing, for example, that error rate is lower for more frequent nouns. If morphological accuracy is impacted by multiword information, then errors should also be lower in more frequent sequences. This is indeed the case (Arnon & Clark, 2011): children's production of irregular plurals in the lab is facilitated in lexically-frequent frames (frames that appear often with the noun in question, like *Brush your-teeth*), and their spontaneous errors are vanishingly rare in such contexts in naturalistic settings. That

is, children's knowledge of irregular forms is modulated by the larger linguistic context. Not taking this into account paints an inaccurate picture of children's morphological knowledge and their ability to use it accurately in production.

Multiword sequences also provide children with correct material to be used in their early productions: in one study, 75% of two-year-olds' early multiword utterances could be derived from previous utterances by using a single combinatorial operation like addition or substitution (Lieven et al., 2003, see also Lieven, Salomo & Tomasello, 2009). More importantly, 50% of the first 400 identifiable multi-word utterances were classified as frozen: there was no evidence that children had productive knowledge of their parts (Lieven et al., 1997). Instead, they seem to rely on frequent multiword combinations used as is. Similar effects can be found at the construction level: children start out using syntactic constructions with verbs that appear frequently in them (e.g., Lieven et al., 1997; Theakston et al., 2004). Construction use is also impacted by multiword frequency: children's early questions draw on lexically-specific phrases (Dąbrowska & Lieven, 2005), and they are more accurate at imitating complex questions when they are more similar to the most prototypical and frequent forms in their input (e.g., *what do you think* is easier to imitate than the less frequent *what do they need*, Dabrowska et al., 2009). Similarly, children show more flexible knowledge of complement clauses when the subject-verb combination is less frequent, making it a better candidate for creating a slot. Four-year-olds were better at shifting from first person to third person for lower frequency subject-verb combinations like *I guess* than for high frequency ones like *I believe* (Brandt, Verhagen, Lieven & Tomasello, 2011). These studies illustrate how multiword information impacts children's ability to generalize and use a range of syntactic constructions.

Until now, we have seen how multiword information can facilitate production and comprehension. However, knowledge of multiword sequences can also hinder accurate production, when the target sequence is in competition with a more frequent one. There are many such examples in the literature. Children's inversion errors in questions can be linked to the frequent non-inverted sequences the question words appear in: children make more errors with sequences that appear more often as non-inverted in their input (Rowland, 2007; Ambridge & Rowland, 2009). A similar link was found for *me-for-I* errors, where children produce incorrect sequences like *me go*: there were more such errors when the sequence appeared often in other, correct, uses (e.g., *Let me go*, (Kirjavainen, Theakston & Lieven, 2009). Once we start looking, quite a few of children's errors can be traced back to competing (correct) multiword sequences (see Theakston & Lieven, 2017).

*Untested predictions.* If lexical and morphological knowledge is tied to larger patterns, then we should see wide-spread effects of multiword information in tasks that assess children's linguistic abilities. Children's ability to produce, inflect and combine words should be facilitated in more frequent sequences across the board, not only for irregular plurals. Children's ability to produce and recognize words should be enhanced by their introduction in frequent frames. In production, this could be tested by showing children pictures of pictures and eliciting naming responses using sequences the words tend to appear in. To give an example, we would compare children's ability to produce a word (e.g., *milk*) following a sequence it appears with often (e.g., *drink your ---*), and following a general question (e.g., *what is this?*). We would expect to see earlier and higher accuracy for the same word when elicited in more frequent sequences (estimated using corpora of child-directed speech). Similar effects are predicted for

comprehension, where the use of frequent frames (used here as broadly as sequences lexical items appear with more often) should enhance lexical recognition. Recent work shows that 6-month-infants can already recognize certain words (earlier than previously thought, (Bergelson & Swingley, 2012). Combining this with our own evidence on infants' sensitivity to multiword frequency (Skarabela et al., 2021), leads to the prediction that infants will recognize more words (or earlier) when embedded in frequent frames. This could be tested using the looking-while listening paradigm (e.g., Fernald & Hurtado, 2006) where infants hear sentences while looking at pictures of two objects. Prompts are usually neutral, using general attention getters like 'look at the dog' or 'where is the doggie?' One could compare gaze patterns in such prompts compared to lexically-specific ones where the word in question appears often in that sequence (e.g., *pet the dog*). Testing the impact of multiword information on the production and recognition of words could reveal earlier or larger lexical knowledge than currently assumed.

The ability to inflect verbs should be similarly affected by multiword information. Knowledge of inflection is often tested by asking children to shift one form into another using general prompts. For instance, having them hear a verb in the present (*Mary walks today*) and asking them to produce it in the past (*What did Mary do yesterday?*). Under the SBH, knowledge of both regular and irregular verb inflection should be impacted by the larger linguistic context: children should have an easier time inflecting words in sequences where the past tense is used more often. These frequencies would have to be extracted from child-directed speech: a first step would be to see whether there are sequences where one tense is used more often than another (e.g., *went to bed* vs. *went to work*). Using frames to re-assess lexical and morphological acquisition could impact when we think certain words are learned. Importantly, further work is needed to determine whether multiword information is similarly facilitative when the word (or morpheme) in question appears before (or in the middle of) the frequent sequence (and not at its end, as has been tested so far).

In this section, I've outlined several ways in which multiword information impacts the learning trajectories of lexical, morphological, and syntactic regularities. Taken together, the findings (which come from researchers not explicitly focused on multiword units) document children's sensitivity to multiword information and the wide-spread effects of this sensitivity on learning a range of linguistic regularities. In the following sections, I take this further to argue that learning from multiword words can facilitate mastery of certain grammatical relations, with consequences for the differential learning trajectories of first and second language learning. These predictions form the core of the Starting Big approach, and distinguish it from other usage-based models.

### **3.2 Multiword units facilitate mastery of semantically opaque grammatical relations**

Larger units (including multiword sequences) are seen as part of linguistic knowledge in many models of linguistic representation. The Starting Big approach takes this a step further to claim that learning from multiword units can actually help children learn semantically opaque relations between words. Languages present us with many such relations in the form of grammatical gender systems, classifier systems, and even verb preposition pairings. Such regularities, which are common cross-linguistically, present learners with a dependency between words that is not semantically transparent. While there are semantic correlates to the choice of form (for instance, event semantics influence the choice of preposition), these correlates are probabilistic, and vary cross-linguistically. Interestingly, children seem to master such semantically opaque

relations relatively easily, while adult second language learners struggle with them, even after extensive exposure: L2 learners have a difficult time learning grammatical relations that are semantically opaque, even when they are fully deterministic (e.g., a noun always belongs to the same gender, see DeKeyser, 2005) for a review)

We can see this differential learning path in the acquisition of grammatical gender and the learning of agreement patterns conditioned on gender. Children master grammatical gender relatively early (see Slobin, 1985 for a cross-linguistic overview), and make few mistakes in spontaneous speech (Bassano, Maillachon & Mottet, 2008; Mariscal, 2009). This is true even in languages that have very elaborate noun class systems, like Bantu languages (Demuth, 1988; Demuth & Ellis, 2009). Adult learners, in contrast, have persistent difficulty with grammatical gender, even after extensive exposure (e.g., Scherag, Demuth, Rösler, Neville & Röder, 2004). Unlike native speakers, adult learners struggle in using gender information predictively: in languages like Spanish, native speakers use the gender marking on articles to predict the upcoming noun (Lew-Williams & Fernald, 2007a), while even proficient non-native speakers do not (Lew-Williams & Fernald, 2007b). That is, native speakers seem to treat the article-noun sequence as a more cohesive unit than do non-native speakers, allowing them to select the correct article in production, and use it to facilitate comprehension.

The Starting Big approach proposes that (some of) this difference is related to the different linguistic building blocks that children and adults draw on, and in particular, to children's greater reliance on multiword units (compared to adults). The idea is that arbitrary relations between words will be learned better when the words in question are initially part of a multiword unit. To take an example, learning that the Spanish noun *pelota* (ball) has to appear with the feminine article *la* will be easier when the two are initially encountered as part of the multiword unit *la-pelota*. Treating the two as one unit, and only then segmenting them, will increase the predictive relations between the article, the noun, and the object it refers to (see Arnon & Ramscar, 2012) for a simulation using discriminative learning). Because of their existing knowledge of words, and the knowledge of specific words in their first language, adults are less likely than children to treat the sequence initially as one unit. This contributes to adults' difficulty in learning dependencies between words.

This proposal makes several concrete predictions about the impact of unit size on learning outcomes. The first is that learning outcomes can be facilitated by manipulating learners' early building blocks: a greater reliance on multiword units will improve learning. The second prediction is that learning from multiword units will facilitate mastery of certain grammatical relations, but not others: learning will not be enhanced when the grammatical element carries independent semantic information (as in the case of plural markers, or marking of natural gender). In these cases, the predictive relation between the grammatical element and the object it modifies is strong enough, and does not need the boost that comes from being initially part of the same unit. These predictions have been supported in a series of artificial language learning studies: children and adults show better learning of gender agreement when exposed to unsegmented input first compared to segmented input first: Importantly, the facilitation is driven by an increased reliance on multiword units (Arnon & Ramscar, 2012; Siegelman & Arnon, 2015; Havron & Arnon, 2020). This facilitative effect was also found when English-speakers were taught a Chinese classifier system (Paul & Grüter, 2016). Importantly, no such facilitation was found when the article carried semantic information, distinguishing between animate and inanimate entities (Siegelman & Arnon, 2015). We are currently investigating the mechanism

underlying these effects using eye-tracking (Abu-Zhaya & Arnon, pre-registered report): if learning from multiword units is facilitative because it increases the association between individual words, we should see increased predictive gaze from the article to the noun when they were initially learned as one multiword unit.

### *Untested predictions*

The Starting Big approach predicts that adults, because of their smaller reliance on multiword units, will be worse than children at learning a host of semantically opaque relations between words, but will struggle less with grammatical elements that carry semantic information. In the real-world, this could translate to a difference between learning case-marking (or tense and number, which carry meaning), and learning grammatical gender (where the division into gender classes is arbitrary for non-natural entities). That is, we would expect adult L2 learners to master case-systems without gender (as in Finnish or Estonian), with more ease than case-systems with gender (as in German), even if the former have more distinct forms to learn. This proposal is not easy to test – since we would have to determine what is meant by mastery (production? Comprehension? Both?), and think how to compare across different language systems (taking into account the complexity and transparency of each of the different case systems). As a first step, we could examine learners' choice of the correct form (using forced-choice trials where two options are heard, but only one is correct) for semantically equivalent cases across languages (e.g., accusative in German, Finnish, Estonian). We would of course have to ensure that participants have the same L1, and that they are similarly proficient in the L2 (both not straightforward to do...).

We can go on to predict that those semantically opaque relations that are better learned from multiword units, will be more prone to simplification by adult L2 learners. This prediction relates to a recent proposal about the impact of L2 learners on morphological complexity. The Linguistic Niche hypothesis proposes a causal link between the proportion of L2 speakers in a community and the degree of morphological complexity of the community's language (Lupyan & Dale, 2010). This link was supported by a large-scale study of over 2000 languages, showing that languages with more L2 speakers have less complex morphology. Fine-tuning this prediction, I propose that the proportion of L2 speakers will have a stronger impact on the morphological complexity of the grammatical relations they struggle with learning. In other words, gender-agreement systems should be more impacted by the proportion of L2 speakers than case-marking systems.

### *3.3 Prior knowledge and experience impact the reliance on multiword units*

I've proposed that learners differ in their reliance on multiword units, and that this reflects various aspects of their prior linguistic experience. One such factor is knowing that words exist as units of representation, and having an existing lexicon. This is what separates infants learning a first language from children and adults learning a second: the former will extract more multiword units, and rely on them more during learning than the latter. However, as detailed above, novel speech input can be perceived as more or less segmented, even after substantial language experience. One important factor contributing to this is the acquisition of literacy. Learning to read doesn't merely add a written dimension to our existing linguistic representations; it can also change those representations. For example, children's awareness of roots is enhanced when they learn to read (e.g., Ravid, 2001). Literacy acquisition also influences speakers' treatment of

words as the relevant unit of processing (Karmiloff-Smith, Grant, Sims, Jones & Cuckle, 1996): the separation between words becomes more salient when speakers are exposed to written text, where words are often separated with spaces.

Accordingly, the Starting Big approach predicts that learning to read will reduce learners' reliance on multiword units, with implications for learning outcomes. Counterintuitively, literacy acquisition is predicted to detrimentally impact the mastery of certain grammatical relations. In line with this, literacy impacts the degree to which learners extract multiword units from novel speech: literate adults show increased reliance on word units compared to illiterate adults (Havron & Arnon, 2016), as do literate children compared to pre-literate ones (Havron & Arnon, 2017). Literacy also impacts learning biases in the expected direction: pre-literate children showed better learning of article-noun agreement patterns compared to mastery of individual nouns, while literate children and adults showed the opposite pattern (Havron, Raviv & Arnon, 2018; Havron & Arnon, 2020). That is, literacy contributes to a reduced reliance on multiword units in learning a novel language.

### *Untested predictions*

So far, the impact of literacy on learning grammatical relations has been studied using artificial languages. One open question is whether this translates to real-world teaching situations: are grammatical gender, and other similar relations, learned better from non-written input? Finding they are would have far-reaching practical implications for second language pedagogy. One way to test this is to compare L2 learning of the same input with and without text, and with and without the text divided into individual words. Looking back at the case of article-noun grammatical gender agreement, one could compare three learning conditions: (a) only auditory where participants see objects and hear their names (including the article) in the novel language, (b) auditory + written-unsegmented where participants see objects, hear their names (including the article) in the novel language, and see an accompanying written text where the article and noun are written as one word, and (c) auditory + written-segmented where participant see objects, hear their names, and see written text where the article and noun are separated into two words. I would predict worse learning in condition (b) compared to the other two (with a possible advantage for the only auditory condition). Other factors, beyond age and literacy, may contribute to learners' tendency to rely on multiword units. In particular, individuals seem to vary in their chunking tendencies (Isbilen, Mccauley, Kidd & Christiansen, 2020), and this may be predictive of second language learning outcomes (Culbertson, Andersen & Christiansen, 2020).

## **4. Multiword or multi-morphemic? Expanding cross-linguistic coverage**

The facilitative role of multiword building blocks stems from the fact that the unit contains the relation to be learned. In the case of grammatical gender agreement, the multiword unit is the article-noun sequence. However, the term multiword unit may be misleading: what seems more relevant is whether the larger unit includes the relation to be learned. In languages like English, where many grammatical relations hold between words, this translates into multiword units. In more morphologically complex languages, however, the relevant 'larger' unit may be the word. This may be especially relevant for agglutinating and polysynthetic languages, where words are polymorphemic (consist of many morphemes), and many of the grammatical relations children need to learn are contained within what is often defined as one

word<sup>1</sup>. Extending the Starting Big approach to such languages predicts that children will represent multimorphemic words alongside their parts; that such representations are also found in the adult lexicon (and influence processing); that children use multimorphemic units in early learning, and that drawing on them can facilitate learning of the grammatical relations conveyed by the individual morphemes and their relation to one another.

There is less work on the acquisition and processing of agglutinating and polysynthetic languages than on isolating languages, or languages with limited inflectional morphology, like English. But the available evidence provides support for these predictions<sup>2</sup>. A recent review summarizes the developmental research on the acquisition of morphology in polysynthetic languages and other highly morphologically complex languages (Kelly, Wigglesworth, Nordlinger & Blythe, 2014). Several interesting generalizations emerge from this literature. Children's early productions in polysynthetic languages are usually monosyllabic, and consequently mono-morphemic (Quechua – Pye, 1992; Navajo – Courtney & Saville-Troike, 2002). For example, at 1;1, a Navajo-learning child produced 'da' (sit) instead of the adult form *ní-d'aah* 'you sit'. Prima facie, this could be taken as evidence against the role of multi-morphemic units in learning. However, as argued above for the word level, the units of production are not necessarily the units of perception: producing a mono-morphemic syllable does not mean the child is not drawing on a larger, multi-morphemic representation. Supporting this, once children start to produce longer sequences, their early multi-morphemic uses are accurate, complex, and contain under-analyzed multi-morphemic 'chunks'. This suggests that the tendency to produce single words/morphemes is driven by production pressures and may not reflect the units of perception. This is somewhat supported by children's error patterns: young Navajo and Quechua learners, for example, rarely make errors in morpheme sequencing, even though older learners do (Courtney and Saville-Troike, 2002); a reflection, perhaps, of 'chunked' multi-morphemic representations. In Inuktitut, a polysynthetic language, children's learning of morphological causation seems to start with the use of unanalysed multi-morphemic chunks (Allen, 1998). That is, children's initial mono-morphemic productions do not preclude the extraction and representation of larger multi-morphemic representations.

In line with this prediction, children acquiring polysynthetic and agglutinating languages use complex inflectional patterns correctly even in their early productions (Allen, 2017). Children learning West Greenlandic, a highly polysynthetic language with numerous derivational affixes, produce many of these affixes correctly by the age of two (Fortesque and Lennert Olsen, 1992). Similarly, two-year-olds acquiring Tamil (an agglutinating language) accurately mark tense, aspect, modality, person, number and gender on the verbs they produce from the start (Raghavendra & Leonard, 1989), expressing more complex relations than found in the early uses of verbs in English-speaking children. Children master highly complex inflectional

<sup>1</sup>How to define a word, and how relevant words are as units cross-linguistically, are both highly debated topics (e.g., Bickel & Zuniga, 2017). The focus on words as the basic unit of meaning is clearly also impacted by the languages we, as researchers, tend to speak, and the writing systems we employ.

<sup>2</sup>Due to space limitations, I only review part of this important literature here (see also Slobin, 1985–1997). I omit the discussion of studies attempting to compare the stages or rate of morphological acquisition across typologically different languages (e.g., Dressler, 2005).

systems early on: Sesotho, an agglutinative Bantu language, has multiple noun classes, which condition agreement with various elements in the sentence, as well as grammatical marking of functions such as passive, applicative, causative, and reflexive on the verb. Amazingly, by 2;6, children are using these markers correctly, and productively (Demuth, 1988). This early mastery is also seen in Chintang, a Sino-Tibetan polysynthetic language, which has very complex verbal inflection paradigms (Stoll et al., 2012): even though Chintang-speaking children are exposed to many more unique verb forms than English-speaking children (1559 vs. 23,811), meaning they hear each less often, they produce a wide and variable range of forms from early on (Stoll, Mazara & Bickel, 2017). By age 2;0, one of the recorded children produced 137 unique verb forms, while by age 3;6 another child produced over 1800 unique verb forms. A reliance on multi-morphemic units could explain how children master morphologically complex systems with relative ease and why they produce correct multi-morphemic sequences early on.

### *Untested predictions*

Much work is needed to extend the Starting Big approach to such languages, and test the predictions that children extract multi-morphemic units, and that such units can facilitate learning of certain grammatical relations between them. One clear prediction is that learning will be impacted by multi-morpheme frequency, such that children will produce morphemes more accurately when they are part of a frequent multi-morphemic sequence. For example, accuracy of verbal inflection in Chintang may be related to the frequency of the larger multi-morphemic sequence. This seems to be the case, at least for the combination of two morphemes: Japanese-speaking children were more accurate to produce past-tense verbs when those appeared with more past-biased verbs (Tatsumi, Ambridge & Pine, 2018). A second prediction is that native adult speakers will also be sensitive to multi-morpheme frequency and show faster processing of higher frequency morphemes: this prediction is hard to evaluate given the scarcity of studies on the processing of polysynthetic and agglutinating languages. A third prediction is that the learning of certain grammatical dependencies between morphemes will be facilitated when starting with multi-morphemic units: here also, the facilitative effect should be dependent on the semantic information conveyed by the individual morphemes. For instance, multi-morphemic units will facilitate learning of the correct matching between affixes and nouns in a semantically opaque noun class system more than for learning the use of affixes marking tense and aspect distinctions. These predictions can be tested using artificial language learning paradigms similar to those used to study the impact of multiword units on learning.

### **Starting Big parallels in other developmental domains**

In this paper, I've focused on the impact of extracting and representing multiword units for language acquisition. More broadly, the Starting Big approach is an account of learning that emphasizes the importance of coarse-to-fine generalizations in development. The basic idea is that order-of-acquisition matters: it is better to learn some generalizations before others. In particular, starting out with broader and less differentiated categories can help learn finer distinctions. So far, our 'coarser' units were multiword ones, and the finer distinctions to be learned were the words themselves, and the relations between them. However, given my domain-general view of language learning, the same kind of logic should apply to other



phenomenon in language, as well as to learning in non-linguistic domains. I briefly mention several other areas in which coarse-to-fine generalizations may facilitate learning. One famous example is the use of chunked board representations by expert chess players, as a way to facilitate memory of individual moves (Gobet & Simon, 1996). A more recent example comes from the domain of face perception. Children treated for congenital cataracts continue to experience difficulty with face-perception, even when the cataract was removed early in life. This difficulty is often explained by a critical period for face processing. A recent paper proposes an alternative explanation (Vogelsang et al., 2018), suggesting the difficulty reflects “the potential downside of high initial visual acuity” (p. 1). Unlike healthy infants, these children did not go through an initial period of low-acuity vision. Instead, once the cataract was removed, their vision had higher acuity than that of newborn infants. The authors propose that the lack of early low-acuity, leads to a reduction in the extended spatial processing (more holistic processing) that is needed for accurate face perception. They support this prediction by showing better face discrimination (and more extended spatial processing) in a convolutional neural network trained first on low-acuity images and only then on high-acuity ones. While the existence of similar processes in children still remains to be shown, this simulation provides additional evidence for the advantage of coarse-to-fine generalizations in learning.

Another domain where degraded input could impact what we learn is our exposure to low-pass filtered speech in the womb. The human foetus is able to hear and process sounds from outside the womb around week 28 (DeCasper and Firth, 1980). The sounds they hear, however, pass through the maternal abdominal wall, which filters out high frequency sounds and enhances low frequency ones. Prosodic information, like stress, is preserved in low-pass filtered speech, but (most) phonetic information is lost. Very speculatively, early exposure to input that is degraded acoustically could enhance learners’ attention to stress (an important cue in many languages for later segmentation). This prediction is compatible with studies of pre-term babies. Healthy pre-term babies who were denied the low-pass experience, are delayed in their learning of prosodic information (Peña, Pittaluga & Mehler, 2010; Ragó, Honbolygó, Róna, Beke & Csépe, 2014), but not phonetic information (Gonzalez-Gomez & Nazzi, 2012). Taken together, these examples aim to highlight the importance of coarse-to-fine generalization during development.

## Summary

We are so used to celebrating words: an infant’s first word is an exciting milestone, the number of words we know is measured throughout life and often taken as an indication of our verbal abilities. In this paper, I’ve highlighted the importance of larger units in understanding what it means to know language, how that knowledge is acquired, and how learning mechanisms are impacted by prior experience.

**Acknowledgments.** I want to thank Shira Tal, Eve V. Clark, Rana Abu-Zhaya and Jenny Culbertson for their helpful comments on previous drafts. All remaining errors are my own. This work was funded by Israeli Science Foundation grant, 445/20

## Bibliography

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage based account of syntactic acquisition. *The Linguistic Review*, 23, 275–290.

- Allen, S. E. M. (1998). Categories within the verb category: learning the causative in Inuktitut. *Linguistics*, 36(4), 633–677.
- Allen, S. E. M. (2017). *Polysynthesis in the acquisition of Inuit languages*. In M. Fortescue, M. Mithun, & N. Evans (eds.), *Handbook of Polysynthesis*. Oxford: Oxford University Press. Retrieved from <https://pdfs.semanticscholar.org/1158/30500571c5942d56785ad37bf746fcd023b6.pdf>
- Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, 42(2), 239–273. <https://doi.org/10.1017/S030500091400049X>
- Ambridge, B., & Rowland, C. (2009). Predicting children's errors with negative questions: Testing a schema-combination account. *Cognitive Linguistics*, 20(2), 225–266. <https://doi.org/10.1515/COGL.2009.014>
- Arnon, I. (2010). Rethinking child difficulty: the effect of NP type on children's processing of relative clauses in Hebrew. *Journal of Child Language*, 37(1), 27–57.
- Arnon, I. (2016). The nature of child-directed speech in Hebrew: frequent frames in a morphologically rich language. In Ruth Berman (ed.), *The Acquisition of Hebrew*. Trends in Language Acquisition Research Series. Amsterdam: John Benjamins.
- Arnon, I., & Christiansen, M. H. (2017). The Role of Multiword Building Blocks in Explaining L1–L2 Differences. *Topics in Cognitive Science*, 9(3), 621–636. <https://doi.org/10.1111/tops.12271>
- Arnon, I., & Clark, E. V. (2011). *Experience, Variation and Generalization: Learning a First Language*. In Inbal Arnon & E. V. Clark, (eds.). Amsterdam: John Benjamins.
- 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.
- 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*, x, xx–xx. <https://doi.org/10.1075/ml.9.3.01arn>
- Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, 92. <https://doi.org/10.1016/j.jml.2016.07.004>
- Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: how order-of-acquisition affects what gets learned. *Cognition*, 122(3), 292–305. <https://doi.org/10.1016/j.cognition.2011.10.009>
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1), 67–82.
- Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the Combinatorial Explosion: An Explanation of n-gram Frequency Effects Based on Naive Discriminative Learning. *Language and Speech*, 56(3), 329–347. <https://doi.org/10.1177/0023830913484896>
- Bannard, C., & Matthews, D. (2008). Stored Word Sequences in Language Learning of Four-Word Combinations. *Psychological Science*, 19(3), 241–248.
- Bassano, D., Maillochon, I., & Mottet, S. (2008). *Noun grammaticalization and determiner use in French children's speech: a gradual development with prosodic and lexical influences*. *Journal of child language* (Vol. 35). <https://doi.org/10.1017/S0305000907008586>
- Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences of the United States of America*, 114(49), 12916–12921. <https://doi.org/10.1073/pnas.1712966114>
- Bergelson, E., & Swingle, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9), 3253–3258. <https://doi.org/10.1073/pnas.1113380109>
- Bickel, B., & Zuniga, F. (2017). The 'word' in polysynthetic languages: Phonological and syntactic challenges. In M. Fortescue, M. Mithun, & N. Evans (eds.), *The Oxford Handbook of Polysynthesis*, Oxford University Press.
- Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *Linguistic Review*, 23(3), 291–320.
- Bod, R. (2009). From exemplar to grammar: a probabilistic analogy-based model of language learning. *Cognitive Science*, 33(5), 752–793. <https://doi.org/10.1111/j.1551-6709.2009.01031.x>
- Bolinger, D. (1968). *Aspects of Language*. New York: Harcourt, Brace & World, Inc.

- Borensztajn, G., Zuidema, W., & Bod, R.** (2009). Children's grammars grow more abstract with age—evidence from an automatic procedure for identifying the productive units of language. *Topics in Cognitive Science*, 1(1), 175–188. <https://doi.org/10.1111/j.1756-8765.2008.01009.x>
- Brandt, S., Verhagen, A., Lieven, E., & Tomasello, M.** (2011). German children's productivity with simple transitive and complement-clause constructions: Testing the effects of frequency and variability. *Cognitive Linguistics*, 22(2), 325–357. <https://doi.org/10.1515/COGL.2011.013>
- Cameron-Faulkner, T., Lieven, E., & Tomasello, M.** (2003). A construction based analysis of child directed speech. *Cognitive Science*, 27(6), 843–873. <https://doi.org/10.1016/j.cogsci.2003.06.001>
- Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A.** (2009). Categorizing words using “frequent frames”: what cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science*, 12(3), 396–406. <https://doi.org/10.1111/j.1467-7687.2009.00825.x>
- Chomsky, N.** (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Clark, E. V.** (1978). Awareness of Language: Some Evidence from what Children Say and Do (pp. 17–43). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-67155-5\\_2](https://doi.org/10.1007/978-3-642-67155-5_2)
- Clark, E. V., & Hecht, B. F.** (1983). Comprehension, Production and Language Acquisition. *Annual Review of Psychology*, 34, 325–349.
- Courtney, E. H., & Saviile-Troike, M.** (2002). Learning to construct verbs in Navajo and Quechua. *Journal of Child Language*, 29(3), 623–654. <https://doi.org/10.1017/S0305000902005160>
- Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J.** (2019). Child-Directed Speech Is Infrequent in a Forager-Farmer Population: A Time Allocation Study. *Child Development*, 90(3), 759–773.
- Culbertson, G., Andersen, E., & Christiansen, M. H.** (2020). EMPIRICAL STUDY Using Utterance Recall to Assess Second Language Proficiency. *Language Learning*, 70, 104–132. <https://doi.org/10.1111/lang.12399>
- Dąbrowska, E., Rowland, C. & Theakston, A.** (2009). The acquisition of questions with long-distance dependencies. *Cognitive Linguistics*, 20(3), 571–597. <https://doi.org/10.1515/COGL.2009.025>
- Dąbrowska, E., & Lieven, E.** (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics*, 16(3), 437–474.
- DeCasper, A. J. & Firth, W. P.** (1980). Of human bonding: newborns prefer their mother's voices. *Science*, 208, 1174–1176.
- DeKeyser, R. M.** (2005). What Makes Learning Second-Language Grammar Difficult? A Review of Issues. *Language Learning*, 55(S1), 1–25. <https://doi.org/10.1111/j.0023-8333.2005.00294.x>
- Demuth, K.** (1988). Noun classes and agreement in Sesotho acquisition. In M. Barlow & C. A. Ferguson (eds.), *Agreement in natural language: approaches, theories and descriptions*. University of Chicago Press.
- Demuth, K.** (2009). Revisiting the Acquisition of the Sesotho Noun Classes. In J. Guo & E. Lieven (eds.), *Crosslinguistic Approaches to the Psychology of Language*. Retrieved from [https://books.google.co.il/books?hl=en&lr=&id=WjZ5AgAAQBAJ&oi=fnd&pg=PA93&dq=Demuth,+1992+Bantu&ots=WR66ttgrST&sig=4b58DO8zOzjRSKthknfKCKVMO1FI&redir\\_esc=y#v=onepage&q=Demuth%2C+1992+Bantu&f=false](https://books.google.co.il/books?hl=en&lr=&id=WjZ5AgAAQBAJ&oi=fnd&pg=PA93&dq=Demuth,+1992+Bantu&ots=WR66ttgrST&sig=4b58DO8zOzjRSKthknfKCKVMO1FI&redir_esc=y#v=onepage&q=Demuth%2C+1992+Bantu&f=false)
- Diessel, H., & Tomasello, M.** (2001). The Development of Relative Clauses in Spontaneous Child Speech. *Cognitive Linguistics*, 11(1–2), 131–151. <https://doi.org/10.1515/cogl.2001.006>
- Dodd, B.** (1975). Children's Understanding of their Own Phonological Forms. *Quarterly Journal of Experimental Psychology*, 27(2), 165–172. <https://doi.org/10.1080/14640747508400477>
- Dressler, W. U.** (2005). Morphological typology and first language acquisition: some mutual challenges. In G. E. Booij, E. Guevara, A. Ralli, S. Sgroi & S. Scalise (eds.), *Proceedings of the fourth Mediterranean Morphology Meeting on Morphology and Linguistic Typology*. University of Bologna.
- Erkelens, M.** (2009). *Learning to categorize verbs and nouns: studies on Dutch*. LOT. Retrieved from <https://www.lotpublications.nl/learning-to-categorize-verbs-and-nouns-learning-to-categorize-verbs-and-nouns-studies-on-dutch>
- Fernald, A., & Hurtado, N.** (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, 9(3), 33–40.
- Fernald, A., Marchman, V. A., & Weisleder, A.** (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, 16(2), 234–248. <https://doi.org/10.1111/desc.12019>
- Fortescue, M., & Lennert Olsen, L.** (1992). The acquisition of West Greenlandic. In D. I. Slobin (ed.), *The Crosslinguistic Study of Language Acquisition* (Vol. 3, pp. 111–220). Hillsdale, NJ: Lawrence Erlbaum.
- Friedmann, N., & Novogrodsky, R.** (2004). The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child Language*, 31(3), 661–681. <https://doi.org/10.1017/S0305000904006269>

- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68(1), 1–76. [https://doi.org/10.1016/S0010-0277\(98\)00034-1](https://doi.org/10.1016/S0010-0277(98)00034-1)
- Gobet, F., & Simon, H. A. (1996). Recall of random and distorted chess positions: Implications for the theory of expertise. *Memory and Cognition*, 24(4), 493–503. <https://doi.org/10.3758/BF03200937>
- Goldberg, A. (2006). *Constructions at Work*. Oxford: Oxford University Press.
- Gonzalez-Gomez, N., & Nazzi, T. (2012). Phonotactic acquisition in healthy preterm infants. *Developmental Science*, 15(6), 885–894. <https://doi.org/10.1111/j.1467-7687.2012.01186.x>
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.
- Havron, N., & Arnon, I. (2016). Reading between the words: The effect of literacy on second language lexical segmentation. *Applied Psycholinguistics*, 1–27. <https://doi.org/10.1017/S0142716416000138>
- Havron, N., & Arnon, I. (2017). Minding the gaps: literacy enhances lexical segmentation in children learning to read\*. *Journal of Child Language*, 1–23. <https://doi.org/10.1017/S0305000916000623>
- Havron, N., & Arnon, I. (2020). Starting Big: The Effect of Unit Size on Language Learning in Children and Adults. *Journal of Child Language*, 48(2), 244–260. doi:10.1017/S0305000920000264
- Havron, N., Raviv, L., & Arnon, I. (2018). Literate and pre-literate children show different learning patterns in an artificial language learning task. *Journal of Cultural Cognitive Science*, 2, 21–33.
- Hyams, N. (1988, August). A Principles-and-Parameters Approach to the Study of Child Language. Retrieved from <https://eric.ed.gov/?id=ED302076>
- Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically Induced Chunking Recall: A Memory-Based Approach to Statistical Learning. *Cognitive Science*, 44, 12848. <https://doi.org/10.1111/cogs.12848>
- Johnson, T., & Arnon, I. (2019). Processing Non-Concatenative Morphology – A Developmental Computational Model. *Proceedings of the Society for Computation in Linguistics*, Vol. 2, Article 61. <https://doi.org/10.7275/gx15-qk46>
- Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2013). Meaning overrides frequency in idiomatic and compositional multiword chunks. In N. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 692–697). Austin, TX: Cognitive Science Society.
- Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2020). Meaningfulness beats frequency in multiword chunk processing. *Cognitive Science*, 44, e12885.
- Juhász, B. J. (2005). Age-of-Acquisition Effects in Word and Picture Identification. *Psychological Bulletin*, 131(5), 684–712. <https://doi.org/10.1037/0033-2909.131.5.684>
- Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences*, 3(9), 323–328. [https://doi.org/10.1016/S1364-6613\(99\)01363-7](https://doi.org/10.1016/S1364-6613(99)01363-7)
- Karmiloff-Smith, A., Grant, J., Sims, K., Jones, M.-C., & Cuckle, P. (1996). Rethinking metalinguistic awareness: representing and accessing knowledge about what counts as a word. *Cognition*, 58(2), 197–219. [https://doi.org/10.1016/0010-0277\(95\)00680-X](https://doi.org/10.1016/0010-0277(95)00680-X)
- Kelly, B., Wigglesworth, G., Nordlinger, R., & Blythe, J. (2014). The acquisition of polysynthetic languages. *Linguistics and Language Compass*, 8(2), 51–64. <https://doi.org/10.1111/lncl.12062>
- Kidd, E., & Bavin, E. L. (2002). English-speaking children’s comprehension of relative clauses: Evidence for general-cognitive and language-specific constraints on development. *Journal of Psycholinguistic Research*, 31(6), 599–617. <https://doi.org/10.1023/A:1021265021141>
- Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children’s processing of relative clauses. *Language and Cognitive Processes*, 22(6), 860–897. <https://doi.org/10.1080/01690960601155284>
- Kirjavainen, M., Theakston, A., & Lieven, E. (2009). Can input explain children’s me-for-I errors? *Journal of Child Language*, 36(5), 1091–1114. <https://doi.org/10.1017/S0305000909009350>
- Lew-Williams, C., & Fernald, A. (2007a). How First and Second Language Learners Use Predictive Cues in Online Sentence Interpretation in Spanish and English. In H. Caunt-Nulton, S. Kulatilake, & I. Woo (eds.), *Proceedings of the 31st Annual Boston University Conference on Language Development* (pp. 382–393). Somerville: Cascadilla Press. Retrieved from <https://pdfs.semanticscholar.org/55b5/3a47314c300cd94cd8d270d11239df2bfff.pdf>
- Lew-Williams, C., & Fernald, A. (2007b). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18(3), 193–198. <https://doi.org/10.1111/j.1467-9280.2007.01871.x>

- Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: a usage-based approach. *Journal of Child Language*, 30(2), 333–370. <https://doi.org/10.1017/S0305000903005592>
- Lieven, E., & Tomasello, M. (2008). Children's first language acquisition from a usage-based perspective. In P. Robinson & N. C. Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 168–196). New York and London: Routledge.
- Lieven, E. V., Salomo, D., & Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, 20, 481–507.
- 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. <https://doi.org/10.1017/S0305000996002930>
- Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, 5(1). <https://doi.org/10.1371/journal.pone.0008559>
- Mahr, T., & Edwards, J. (2018). Using language input and lexical processing to predict vocabulary size. *Developmental Science*, 21(6), e12685.
- Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularisation in Language Acquisition. *Monographs of the Society for Research in Child Development*, 57(4), 1–178.
- Mariscal, S. (2009). Early acquisition of gender agreement in the Spanish noun phrase: starting small. *Journal of Child Language*, 36(1), 143–171. <https://doi.org/10.1017/S0305000908008908>
- Maslen, R., Theakston, A. L., Lieven, E., & Tomasello, M. (2004). A Dense Corpus Study of Past Tense and Plural Overregularization in English. *Journal of Speech, Language and Hearing Research*, 47, 1319–1333.
- Matthews, D. E., & Theakston, A. L. (2006). Errors of Omission in English-Speaking Children's Production of Plurals and the Past Tense: The Effects of Frequency, Phonology, and Competition. *Cognitive Science*, 30, 1027–1052. <https://doi.org/10.1207/s15516709cog000>
- McCauley, S. M., & Christiansen, M. H. (2017). Computational Investigations of Multiword Chunks in Language Learning. *Topics in Cognitive Science*, 9(3), 637–652. <https://doi.org/10.1111/tops.12258>
- McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, 126(1), 1–51. <https://doi.org/10.1037/rev0000126>
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1), 91–117. [https://doi.org/10.1016/S0010-0277\(03\)00140-9](https://doi.org/10.1016/S0010-0277(03)00140-9)
- Moran, S., Blasi, D. E., Schikowski, R., Küntay, A. C., Pfeiler, B., Allen, S., & Stoll, S. (2018). A universal cue for grammatical categories in the input to children: Frequent frames. *Cognition*, 175, 131–140. <https://doi.org/10.1016/j.COGNITION.2018.02.005>
- Noble, K. G., Farah, M. J., & McCandliss, B. D. (2006). Socioeconomic background modulates cognition–achievement relationships in reading. *Cognitive Development*, 21(3), 349–368. <https://doi.org/10.1016/j.COGDEV.2006.01.007>
- Paul, J. Z., & Grütter, T. (2016). Blocking Effects in the Learning of Chinese Classifiers. *Language Learning*, 66(4), 972–999. <https://doi.org/10.1111/lang.12197>
- Peña, M., Pittaluga, E., & Mehler, J. (2010). Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences of the United States of America*, 107(8), 3823–3828. <https://doi.org/10.1073/pnas.0914326107>
- Peters, A. M. (1977). Language Learning Strategies: Does the Whole Equal the Sum of the Parts? *Language*, 53(3), 560–573.
- Peters, A. N. N. M. (1983). *The units of language acquisition*. Cambridge: Cambridge University Press.
- Pinker, S. (1999). *Words and rules: The ingredients of language*.
- Pye, C. (1992). The acquisition of K'iche' Maya. In D. I. Slobin (ed.), *The Crosslinguistic Study of Language Acquisition* (Vol. 3, pp. 221–308). Hillsdale, NJ: Lawrence Erlbaum.
- Ravid, D. (2001). Learning to spell in Hebrew: Phonological and morphological factors. *Reading and Writing*, 14, 459–485. <https://doi.org/10.1023/A:1011192806656>
- Raghavendra, P., & Leonard, L. (1989). The acquisition of agglutinating languages: Converging evidence from Tamil. *Journal of Child Language*, 16(2), 313–322. doi:10.1017/S0305000900010436
- Ragó, A., Honbolygó, F., Róna, Z., Beke, A., & Csépe, V. (2014). Effect of maturation on suprasegmental speech processing in full- and preterm infants: A mismatch negativity study. *Research in Developmental Disabilities*, 35(1), 192–202. <https://doi.org/10.1016/j.ridd.2013.10.006>
- Real, F., & Christiansen, M. H. (2007a). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57(1), 1–23. <https://doi.org/10.1016/j.JML.2006.08.014>

- Real, F., & Christiansen, M. H. (2007b). Word chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology* (2006), 60(2), 161–170. <https://doi.org/10.1080/17470210600971469>
- Rowland, C. F. (2007). Explaining errors in children's questions. *Cognition*, 104(1), 106–134. <https://doi.org/10.1016/J.COGNITION.2006.05.011>
- Rumelhart, D. E., & McClelland, J. L. (eds) (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. 1). Cambridge, MA: MIT Press.
- Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. *Annual Review of Psychology*, 69(1). <https://doi.org/10.1146/annurev-psych-122216-011805>
- Scherag, A., Demuth, L., Rösler, F., Neville, H. J., & Röder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition*, 93(3), B97–108. <https://doi.org/10.1016/j.cognition.2004.02.003>
- Siegelman, N., & Arnon, I. (2015). The advantage of starting big: Learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *Journal of Memory and Language*, 85, 60–75. <https://doi.org/10.1016/j.jml.2015.07.003>
- Slobin, D. I. (1985). *The Crosslinguistic Study of Language Acquisition: The Data* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum.
- Siyanova-Chanturia, A., & Martinez, R. (2014). The Idiom Principle Revisited. *Applied Linguistics*, 36(5), amt054. <https://doi.org/10.1093/applin/amt054>
- Skarabela, B., Ota, M., O'Connor, R., & Arnon, I. (2021). 'Clap your hands' or 'take your hands'? One-year-olds distinguish between frequent and infrequent multiword phrases. *Cognition*, 211, 104612. <https://doi.org/10.1016/j.cognition.2021.104612>
- Soderstrom, M. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, 49(2), 249–267. [https://doi.org/10.1016/S0749-596X\(03\)00024-X](https://doi.org/10.1016/S0749-596X(03)00024-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. <https://doi.org/10.1111/j.1551-6709.2008.01004.x>
- Stoll, S., Bickel, B., Lieven, E., Paudyal, N. P., Banjade, G., Bhatta, T. M., Gaenszle, M., Pettigrew, J., Ray, I. P., Ray, M., & Ray, N. K. (2012). Nouns and verbs in Chintang: Children's usage and surrounding adult speech. *Journal of Child Language*, 39, 284–321. doi: 10.1017/S0305000911000080
- Stoll, S., Mazara, J., & Bickel, B. (2017). The acquisition of polysynthetic verb forms in Chintang. (August 2016).
- Tatsumi, T., Ambridge, B., & Pine, J. M. (2018). Disentangling effects of input frequency and morphophonological complexity on children's acquisition of verb inflection: An elicited production study of Japanese. *Cognitive Science*, 42 (Suppl 2), 555–577. <https://doi.org/10.1111/cogs.12554>
- Theakston, A., & Lieven, E. (2017). Multiunit Sequences in First Language Acquisition. *Topics in Cognitive Science*, 9(3), 588–603. <https://doi.org/10.1111/tops.12268>
- Theakston, A. L., Lieven, E. V., Pine, J. M., & Rowland, C. F. (2004). Semantic generality, input frequency and the acquisition of syntax. *Journal of Child Language*, 31(1), 61–99. doi:10.1017/S0305000903005956.
- Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition*.
- 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. <https://doi.org/10.1111/j.1467-9922.2010.00622.x>
- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92(1–2), 231–270. <https://doi.org/10.1016/j.cognition.2003.10.008>
- Vogelsang, L., Gilad-Gutnick, S., Ehrenberg, E., Yonas, A., Diamond, S., Held, R., & Sinha, P. (2018). Potential downside of high initial visual acuity. *Proceedings of the National Academy of Sciences of the United States of America*, 115(44), 11333–11338. <https://doi.org/10.1073/pnas.1800901115>
- Wang, H., Höhle, B., Ketrez, F. N., & Küntay, A. C. (2011). Cross-Linguistic Distributional Analyses with Frequent Frames: the Cases of German and Turkish, 628–640.
- Weisleder, A., & Waxman, S. R. (2010). What's in the input? Frequent frames in child-directed speech offer distributional cues to grammatical categories in Spanish and English. *Journal of Child Language*, 37(5), 1089–1108. <https://doi.org/10.1017/S0305000909990067>