This is a contribution from *Experience, Variation and Generalization. Learning a first language*.
Edited by Inbal Arnon and Eve V. Clark.

Units of learning in language acquisition

Inbal Arnon
University of Haifa

Textbook descriptions of how production develops in first language acquisition often move from babbling (producing syllables), through single-word utterances, to multi-word combinations. This progression emphasizes the small-to-big aspect of language learning where each stage involves larger, more structured linguistic units. In this chapter, I discuss an equally important, but often neglected process: the move from large unanalyzed units to the identification of smaller more structured ones. I present evidence for such Gestalt processes in a range of linguistic domains (phonetics, morphology, and syntax), and suggest that they play an important role in first language learning by offering children another route into linguistic structure. I discuss implications for identifying early building blocks for language learning.

Keywords: Gestalt processes; multi-word phrases; units of learning

1. Introduction

Textbook descriptions of how production develops in first language acquisition often move from babbling (producing syllables), through single-word utterances, to multi-word combinations and eventually the production of complex sentences (e.g. Bloom 1994; Clark 2009; Goodluck 1991). This progression is described in terms of moves through consecutive stages, each involving larger, more structured linguistic units, starting with syllables, then moving on to words, to word-combinations, and then to constructions. Such accounts emphasize the combinatorial, or small-to-large, direction of learning, and hence the ability to create larger and more complex units from smaller ones – combining syllables into words, deriving morphologically complex words from words plus grammatical morphemes, and generating larger phrases and sentences from combinations of single words.

Such accounts capture the 'generative' aspect of language. But they mask or ignore what may be an equally important process: the move from large unanalyzed units to the identification and analysis of smaller more structured ones. I will refer to this class of processes as Gestalt processes (after Peters 1977), or big-to-small,

for short. I use this expression for processes where the child learns from an initially unanalyzed whole about the parts inside it, for example, using whole words to identify phonetic contrasts between the segments of a language (Lindblom 1992), or learning from a formulaic chunk like *how-are-you* about the words it is made up from. Such processes are similar to combinatorial ones in that the product of learning is a more complex, structured representation. But they are different in the units they operate over. In short, children can learn from a range of units, not just from atomic ones. Their own early units in production do not necessarily correspond to adult units, nor to linguists' units of analysis.

The significance of Gestalt processes in language learning is highlighted by the continuous nature of speech. Infants don't hear adult speech neatly separated into phonemes, morphemes, and words. To discover these units, children first need to break into the speech stream and identify (some of) the relevant units in their language (e.g. Goldinger & Azuma 2003; Kuhl 2004). This process necessarily involves decomposing large chunks – stretches of continuous, un-segmented speech – into smaller chunks or units such as syllables, words, and morphemes. Early speech segmentation, one of the first tasks infants are faced with, involves a Gestalt process. Yet Gestalt processes in language learning have gone largely unstudied.

One reason is that such processes fall by the wayside in theories that emphasize the rule-governed aspects of language. When linguistic knowledge is viewed as consisting of atomic elements and rules to combine them (e.g. Chomsky 1965; Pinker 1999; Jackendoff 2002), then these are also the elements that researchers look for in children's language. A words-and-rules characterization of adult language tends to focus on (a) combinatorial properties of child speech, and (b) adult-like units – phonemes, morphemes, words – in children's speech. Theories of grammar and development that emphasize the centrality of rules in language (e.g. Borer & Wexler 1987; Pinker 1991) have led to a focus on the combinatorial aspect of children's language learning.

However, more recent theories of grammar acknowledge the role of larger patterns, like constructions (Croft 2001; Fillmore 1988; Goldberg 1995, 2006, Kay & Fillmore 1999; Pollard & Sag 1994). Linguists now view a larger set of units as potential building blocks for language (e.g. Bybee 1998, 2001). Within developmental research, there is also increasing support for usage-based theories of learning (Tomasello 2003) where the speech adults address to children plays a crucial role, and where abstract categories and rules are derived from larger stored chunks of language. These views of language structure and learning provide a better theoretical anchor for understanding Gestalt processes, and offer more incentive to seek them out for study.

2. Gestalt processes in learning

Evidence for Gestalt processes is apparent at many levels of linguistic analysis. Segmenting words from speech necessarily involves a big-to-small process. How infants achieve this task has been the focus of much research (e.g. Saffran et al. 1996; Jusczyk 1999). For example, Saffran and her colleagues suggested that one way infants solve this problem is by using the transitional probabilities between segments and syllables as a cue to boundaries for recurring sequences, and ultimately then for word segmentation. The transitional probabilities from one sound segment to the next within words are higher than those between words. While transitional probabilities are not readily described in Gestalt terms, their use in language analysis does reflect a Gestalt process. For transitional probabilities to be useful, they must be calculated over strings of consecutive words. This is applied implicitly in studies that test the human ability to make use of such information when infants, children, and adults are always exposed to unsegmented sequences of words (Saffran 2003).

Peters (1977) described one child's use of what I will call the *whole-utterance path* to identifying words. Instead of going from single words to longer utterances, this child's early attempts at communicating verbally consisted of adult-like prosodic contours that only later became recognizable as distinct words. For instance, at 1;7 the child produced the sequence [a lər ri go mu nay] (interpreted by Peters as 'I like read good moon night') while he was holding out the book *Good Night Moon*. Only later did the child also produce the individual words here in other utterances.

The idea of whole-word phonology, first formulated by Ferguson & Farwell (1975; see also Macken 1979; Waterson 1971), is another instance of a Gestalt approach to learning. Here, lexical words provide the basis for children's early learning of the segmental inventory of their language. Children start out with whole word forms from which they extract information about relevant phonological contrasts. As their grasp of the phonological system grows, their production of early words may become temporarily less accurate. Ferguson & Farwell, for example, looked at the developmental trajectory of English *pretty* in one child. The child first produced this word at 10 months in "almost perfect phonetic form" (1975:432), at a stage where most of his other words were monosyllabic and lacked all consonant clusters. At 1;9, as he built up phonological classes and patterns, his phonetic form for *pretty* changed to [piti], less accurate but a better reflection of the state of his phonological system at the time, with CV syllables dominant. This pattern of early phonetic accuracy followed by less accurate but internally consistent forms in children's words has been observed in other studies too (e.g. McCune & Vihman 2001).

Vihman and her colleagues have also emphasized the big-to-small direction of phonological acquisition with children's reliance on phonological templates (Velleman & Vihman 2002; Vihman 1996). Children appear to extract templates from words they acquired early on. These then serve as a filter through which children acquire new words. In fact, they often appear to gravitate towards adult words that fit their existing templates.

In a recent study, Swingley (2005) suggested that children can also learn about the dominant stress patterns of a language from their own early word forms. By eight months, English-speaking infants exhibit a trochaic bias in speech segmentation. They prefer to group together syllables that have a trochaic or strong-weak stress pattern rather than syllables with an iambic or weak-strong pattern. How does this preference emerge? Swingley suggested that infants accumulate potential word-forms by grouping together clusters of syllables that occur together often and have a high transitional probability. He used a computational model to show that most potential word-forms have trochaic stress. Infants can therefore use early word forms to extract a dominant stress pattern that they can then use in the segmentation of novel words. Here too, then, children extract information from early, unanalyzed forms.

The view that children learn from unanalyzed wholes is also found in morphological acquisition. Children may learn inflections in a rote-based fashion. They start with inflected wholes and only later learn to identify the morphological components. In support of this idea, Wilson (2003) showed how the frequency of copula omission errors in English (as in *he sleeping*) varies according to the subject-type. Omission was more likely with lexical NPs than with pronominal forms used as subjects, and more likely with lower frequency than with higher frequency pronouns. Children's acquisition of subject-verb agreement in Portuguese (Rubino & Pine 1998) shows a similar pattern. Omission rates vary with the number/person of the subject nominal, going from zero errors for the more frequent first-person subjects to 43% errors with the less frequent third-person-plural subjects. These findings then illustrate a whole-word path to morphological acquisition.

Usage-based studies also suggest a whole-chunk path to construction learning. They document the item-based nature of children's early utterances, compared to later, more abstract and productive knowledge of the relevant structures (Tomasello 2000). For example, in transitive constructions, children seem to start with verb-specific uses and representations before they develop any verb-general ones (e.g. Abbot-Smith, Lieven & Tomasello 2007; Akhtar 1999; Tomasello, Brooks, & Stern 1998). Similar arguments have been made for the acquisition of relative clauses (e.g. Diessel & Tomasello 2000) and passives (Tomasello et al. 1998). Item-based productions appear not only with specific verbs, but also with specific verb-noun combination, indicating the reliance on unanalyzed chunks: many early

transitive utterances have the form *I'm verb-ing it* (Lieven, Pine & Rowland 1998). In Gestalt terms, children's construction learning seems to involve a big-to-small process: They move from unanalyzed chunks to more abstract constructions.

These findings demonstrate that Gestalt processes show up in learning at many levels of linguistic analysis. We can identify Big-to-Small processes in learning about phonemes, morphemes, words, and constructions. A good deal of learning in language can be described as a move from large, unanalyzed units to small, more structured ones.

3. The current findings

The three chapters in this section have highlighted two different aspects of Gestalt learning of linguistic structure. Vihman & Vihman discuss how whole-forms are used to extract phonological structure, while both Brandt and Bannard & Mathews emphasize the importance of variation in extracting regularities from less analyzed or unanalyzed units.

Using corpus data on the development of one child, Vihman & Vihman track the development of segmental phonology. They illustrate the use of templates in early production, and show how these templates (which differ from child to child) serve as a skeleton for selecting and adapting adult words. Their work emphasizes the role of whole-words in discovering phonetic units. Under this approach, lexical words form the basis for building an early segmental inventory. The contrasts that are acquired first influence the new words the child will learn.

In her chapter, Brandt offers a whole-chunk approach to the discovery of grammatical roles. She studies the gradual development of a productive SUBJECT category in German. Her work, combining corpus analysis and experimental manipulations, documents the heightened chunk-like nature of subject-verb combinations that (1) co-occur frequently and (2) don't often appear with other elements. The findings document the existence of under-analyzed units in children's speech, and show how pragmatic information – in this case the diversity of conversational situations tied to specific verb uses–-influences the generalizations that children make.

Bannard & Matthews investigate factors that might guide Gestalt learning. Under the assumption that children's early units are concrete and less abstract than the adult's, what factors influence or drive generalization? They show that children are more willing to form generalizations, shown by their extraction of slots from frames, when there is a lot of variability in the final slot (high slot entropy), and when the words appearing in a given slot are similar to each other in meaning (low semantic density).

All three chapters demonstrate how taking Big-to-Small processes into account can further our understanding of first language development. In looking at phonological development, Gestalt processes offer an explanation for the higher phonetic accuracy of early word attempts compared to later ones (Vihman & Vihman). In the realm of syntactic processing, the focus on big-to-small processes can both account for children's increased difficulty in using certain sequences productively (Brandt), and allow us to investigate the factors that enable children to segment certain early units and then use them in making early syntactic generalizations (Bannard & Matthews).

4. Implications

I have focused on Gestalt processes in this chapter not because they are more important to the acquisition process than combinatorial (small-to-big) processes, but rather because even though they are equally important, they have been traditionally under-studied. Children's speech does not always fit a small-to-big trajectory. They produce many single-word utterances, and, at the same time, attempt to produce multi-word ones (Clark 1974; Peters 1983). Their early word-productions are sometimes more accurate than their later productions, phonologically and morphologically (Ferguson & Farwell 1975), and some of their early utterances appear to present complexity that is not found in the rest of their speech (for instance by showing correct agreement patterns, Peters 1983).

Children's simultaneous use of more advanced and less advanced forms, and the fact that, of necessity, they are segmenting speech and discovering structure at the same time, suggest that Gestalt and analytical processes are complementary. Acknowledging the role of big-to-small processes in the learning process raises a new set of questions and challenges, as well as adding new ways of accounting for children's variable performance. I outline some of these below.

How can we identify the larger units involved in Gestalt processes? One major challenge in studying Gestalt processes, especially ones leading to the extraction of morphological and syntactic structure, lies in identifying the larger units they apply to. By the time children start to talk, they have already done a lot of analysis and segmentation. Moreover, what they produce may be different from what they comprehend or learn from (Clark & Hecht 1983). Just as children's truncated word forms don't mean that they lack a representation for the whole word (e.g. Dodd 1975), their use of single-word utterances doesn't mean they don't have representations in memory of larger units as well. The fact that children's early speech consists mostly of single word utterances doesn't mean they aren't using larger multi-word chunks in the learning process.

One way to identify potential input units for Gestalt processes is to analyze the speech of young children. In a series of studies, Lieven and her colleagues have explored ways of assessing productivity in children's early utterances. Using the trace-back method, which involves going back in time to see where the constituent forms found in later utterances could have come from developmentally, they demonstrate that early language isn't as productive as it may seem at first glance (Lieven, Pine, & Baldwin 1997; Lieven et al. 2003; Lieven, Salomo, & Tomasello 2009). Of particular interest here are multi-word utterances that Lieven and her colleagues classified as frozen: in analyzing their use, they found no indication that young children had productive knowledge of their parts. In fact, up to 50% of the first 400 identifiable multi-word utterances produced by children were classified as frozen. Such 'chunks' are good candidates for Gestalt processes since they are not yet fully analyzed.

Recently developed computational tools may offer a more systematic way of identifying Gestalt candidates in children's early productions. Borensztajn, Zuidema and Bod (2009) use data-oriented parsing (Bod 1998), in combination with statistical parsing techniques, to 'identify the most likely primitive units that were used […] to produce the utterances in a given corpus'. Utterances are decomposed into the fragments they were most likely generated from – these fragments are the models' building blocks. This method shows how children's grammars become more abstract with age: early fragments (called *constructions* by Bod and his colleagues) had more lexical nodes than later ones, meaning that they were more lexically-specific than and less abstract than later productions. This method can also be used to uncover multi-word chunks at different stages in development. At any given point in time, the fragments that have no slots (non-terminal nodes), and that contain more than one word, are likely to be stored and used as single units, and so serve as input for Gestalt processes.

Computational models of segmentation offer additional insights. These models were not designed with the goal of identifying multi-word chunks. But, in using different cues to segment speech into words (stress, transitional probabilities), they all end up with some 'word' units that consist of more than one word. Such units are often discarded as errors that are insignificant to the learning process, but they may actually provide insight into the nature of children's early units. Swingley (2005) extracted early word-forms by clustering together syllables that score high on frequency and mutual information (how informative syllables are about one another). In his English data, the majority of 'false alarms' (syllable sequences erroneously classified as one word) were frequently co-occurring word sequences like *come on* or *want some more*. He found similar patterns in the Dutch data he analyzed.

Using transitional probabilities as a cue to word boundaries also leads to a number of multi-word 'words' (Goldwater et al. 2009). Between 15% and 30% of

the proposed 'words' (depending on whether words are assumed to be statistically independent or not), are multi-word strings like *look at this*. A similar pattern is found when distributional regularity (frequent co-occurrence in a variety of contexts) and phonotactic constraints are used to segment speech (Brent & Cartwright 1996). At least 10% of the proposed word candidates are classified as 'concatenation errors' where two or more words were not properly segmented.

In all these computational simulations, the proposed lexicon contained a substantial proportion of multi-word 'words'. Whatever segmentation criterion was applied, it led to the treatment of some multi-word sequences as single units. Instead of treating these as undesirable errors, we can use them to generate predictions about the kinds of multiword sequences that are initially more likely to be 'chunked' and undergo Gestalt processes. A closer look at these errors suggests several categories that are especially prone to such 'chunking'. The first is the category of socialized routines: expressions like *byebye*, *thank you*, or *night night*; and positive exclamations like *good girl*! or *clever boy*! Such expressions are unlikely to appear in diverse linguistic or social contexts (we rarely say *thank them*). Determiner-noun sequences are also repeatedly identified as one word: of the 70 most frequent multi-word 'words' in Goldwater et al. (2009), almost 20% consisted of determiner-noun combinations. Interestingly, this pattern aligns with reports that children start out treating determiner and nouns as one unit (e.g. Carroll 1989), and that such treatment is beneficial to the learning process (Arnon & Ramscar 2009). Among other examples are such frequent frames as *do you want*, and labeling frames like *that's a* or *it's a*.

If children use segmentation strategies that are anything like the ones used in the simulations (not an unlikely scenario since these were inspired by behavioral findings), then they too will end up with quite a few multi-word units. The simulations offer further insight into the kinds of sequences that could remain as unanalyzed chunks even after initial segmentation – namely, social routines, determiner-noun combinations, and certain frequent frames. The combined insights from computational corpus studies and models of segmentation offer ways of identifying good candidates for Gestalt processes.

Are Gestalt processes affected by variation? In assigning Gestalt processes a role in the process of language acquisition, we necessarily raise new questions about how larger, less analyzed, units become more structured. What are the necessary and sufficient conditions that allow children to segment and analyze these larger units? The findings discussed here highlight the importance of systematic variation in Gestalt learning processes. Hearing a frame that occurs with many dissimilar nouns leads to more productive knowledge of how that frame is used (Bannard & Matthews). Hearing a verb with different kinds of subjects gives children more flexibility in their own productions, and leads to a deeper understanding of the

grammatical category of subject (Brandt). In both cases, our underlying assumption is that children extract structure by comparing utterances that are similar on certain dimensions (e.g. two phrases starting with *a drink of*).

This leads to further questions. First, how do children discover the *relevant* dimensions for comparison? Second, how important is temporal proximity in this process of comparison? Is it enough that children hear the plural form *cats* one day, and the singular form a month later, or do the forms need to appear within a certain temporal window? Interestingly, child-directed speech seems to be structured (at least in part) in a way that draws attention to similarity between forms in a temporally restricted space. Child-directed speech contains a high proportion of variation-sets: sequences of utterances with partial lexical overlap like *Where is **the bunny**? I'm holding **the bunny*** (e.g. Küntay & Slobin 1996; Waterfall 2006). The presence of such sets correlates with children's syntactic acquisition (Waterfall 2006). The ability to compare utterances close in time appears to facilitate linguistic generalizations, but further research is needed to delimit the conditions under which children can make useful similarity comparisons.

Gestalt processes and individual differences. Taking Gestalt processes seriously can cast additional light on documented developmental patterns. By granting that children's early units are less analyzed, we can explain U-shaped patterns in production where early, apparently accurate forms are displaced by later, erroneous ones that are then replaced by the correct forms. In finding out why early word production becomes less accurate, or why certain early utterances display syntactic and morphological complexity not displayed elsewhere in the child's speech, we can draw on the notion of early holistic units. Forms that seem more developed than the rest of the system likely result from children's reliance on under- or unanalyzed chunks.

The same notion appears useful in the study of individual differences in language learning. If early holistic forms serve as a basis for learning linguistic structure, and if those forms differ from child to child as a function of, among other things, each child's specific linguistic experience, then (1) we expect there should be subtle differences in early generalizations, for instance in the specific phonological contrasts that emerge first (see Vihman & Vihman), and we can look for them; and (2), we can make concrete predictions about such differences based on the input children are exposed to. Elements that co-occur frequently in child-directed speech may be extracted and stored as 'chunks', large building blocks in children's representations in memory, that lead in turn to individual learning trajectories.

One recent demonstration is found in a paper on children's *Me*-for-*I* errors (Kirjavainen, Theakston, & Lieven 2009). Children occasionally make case-marking errors in English, using the accusative-marked pronoun (*me*) where a nominative one (*I*) should have been used instead (as in *me [was] crying*). The

rate of such errors is correlated with the number of uses of *me* in first-person-singular preverbal contexts, as in *Let me do it,* in the speech directed to individual children. Children were more likely to reach an erroneous conclusion about where accusative-marked pronouns appear when the input they were learning from included more sequences where a verb (*do*) was preceded by an accusative pronoun (*me*). In short, when children were more likely to use such chunks in the generalization process, they made more errors.

5. Conclusion

In this chapter I have emphasized the importance of big-to-small processes in language learning. Such processes, while understudied, are not rare. The move from larger less analyzed or unanalyzed units to smaller, more structured ones is found in phonological, morphological and syntactic development. Such a move is consistent with the input children receive: speech doesn't come neatly divided into words, and children's adoption of chunks larger than words would explain why certain forms in their speech seem to be produced more accurately, or seem to be unexpectedly complex early on. Consideration of Gestalt processes emphasizes another route for young children into linguistic structure. This route adds to the range of units children can learn from, and it views the young child slightly less as an 'analyzer', and rather more as a communicator – a communicator that starts with form-meaning pairings that only later get fully analyzed.

Scientific study of Gestalt processes that goes beyond anecdotal evidence faces two major challenges: how to identify the units involved in such processes, and how to document the developmental move from larger to smaller units. I hope that future combinations of behavioral and computational methods will be able to address both challenges and so make Gestalt processes an integral part of our models of language development.

References

Abbot-Smith, K., Lieven, E. & Tomasello, M. 2007. Graded representations in the acquisition of German and English transitive constructions. *Cognitive Development* 23: 48–66.

Arnon, I. & Ramscar, M. 2009. Order-of-acquisition affects what gets learned. In *Proceedings of the 31th Annual Conference of the Cognitive Science Society,* N.A. Taatgen & H. van Rijn (eds), 2112–2117. Cognitive Science Society, Amsterdam.

Akhtar, N. 1999. Acquiring basic word order: Evidence for data-driven learning of syntactic structure. *Journal of Child Language* 26: 339–356.

Bloom, P. 1994. *Language Acquisition: Core Readings*. Cambridge MA: The MIT Press.
Bod, R. 1998. *Beyond Grammar: An Experience-based Theory of Language*. Stanford CA: CSLI.
Borensztajn, G., Zuidema, W. & Bod, R. 2009. Children's grammars grow more abstract with age. *Topics in Cognitive Science* 1: 175–188.
Borer, H. & Wexler, K. 1987. The maturation of syntax. In *Parameter Setting,* T. Roeper & E. Williams (eds), 123–172. Dordrecht: Reidel.
Brent, M.R. & Cartwright, T.A. 1996. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition* 61: 93–125.
Bybee, J.L. 1998. The emergent lexicon. In *Proceedings of the Chicago Linguistic Society* 34: *The Panels,* M.C. Gruber, D. Higgins, K.S. Olson & T. Wysocki (eds), 421–435. Chicago IL: Chicago Linguistic Society.
Bybee, J. 2001. *Phonology and Language Use*. Cambridge: CUP.
Carroll, S. 1989. Second-language acquisition and the computational paradigm. *Language Learning* 39: 535–594.
Chomsky, A.N. 1965. *Aspects of the Theory of Syntax*. Cambridge MA: The MIT Press.
Clark, R. (1974). Performing without competence, *Journal of Child Language*, 1: 1–10.
Clark, E.V. 2009. *First Language Acquisition,* 2nd edn. Cambridge: CUP.
Clark, E.V. & Hecht, B.F. 1983. Comprehension, production, and language acquisition. *Annual Review of Psychology* 34: 325–349.
Croft, W. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: OUP.
Diessel, H. & Tomasello, M. 2000. The development of relative constructions in early child speech. *Cognitive Linguistics* 11: 131–152.
Dodd, B. 1975. Children's understanding of their own phonological forms. *Quarterly Journal of Experimental Psychology* 27: 165–172.
Ferguson, C.A. & Farwell, C.B. 1975. Words and sounds in early language acquisition. *Language* 51: 419–439.
Fillmore, C.J. 1988. The mechanisms of 'Construction Grammar'. In *Proceedings of Berkeley Linguistics Society* 14, 35–55. Berkeley CA.
Goldberg, A.E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago IL: Chicago University Press.
Goldberg, A.E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: OUP.
Goldinger, S.D. & Azuma, T. 2003. Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics* 31: 305–320.
Goldwater, S., Griffiths, T.L. & Johnson, M. 2009. A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition* 112: 21–54.
Goodluck, H. 1991. *Language Acquisition: A Linguistic Introduction*. Oxford: Blackwell.
Jackendoff, R. 2002. *Foundations of Language*. Oxford: OUP.
Jusczyk, P.W. 1999. How infants begin to extract words from speech. *Trends in Cognitive Sciences* 9: 323–328.
Kay, P. & Fillmore, C.J. 1999. Grammatical constructions and linguistic generalizations: The *What's X doing Y?* construction. *Language* 75: 1–33.
Kirjavainen, M.M.M., Theakston, A.L. & Lieven, E.V.M. 2009. Can input explain children's *me*-for-*I* errors? *Journal of Child Language* 36: 1091–1114.

Kuhl, P.K. 2004. Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience* 5: 831–843.
Lieven, E., Pine, J. & Baldwin, G. 1997. Lexically-based learning and early grammatical development. *Journal of Child Language* 24: 187–220.
Lieven, E., Behrens, H., Speares, J. & Tomasello, M. 2003. Early syntactic creativity: A usage-based approach. *Journal of Child Language* 30: 333–370.
Lieven, E., Salomo, D. & Tomasello, M. 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics* 20: 481–507.
Lindblom, B. 1992. Phonological units as adaptive emergents of lexical development. In *Phonological Development: Models, research, implications,* C.A. Ferguson, L. Menn & C. Stoel-Gammon (eds), 131–163. Parkton MD: York Press.
Macken, M.A. 1979. Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua* 49: 11–49.
McCune, L. & Vihman, M.M. 2001. Early phonetic and lexical development: A productivity approach. *Journal of Speech, Language, and Hearing Research* 44: 670–684.
Peters, A.M. 1977. Language learning strategies: Does the whole equal the sum of the parts? *Language* 53: 560–573.
Peters, A.M. 1983. *The Units of Language Acquisition*. Cambridge: CUP.
Pinker, S. 1991. Rules of language. *Science* 253: 530–535.
Pinker, S. 1999. *Words and Rules: The Ingredients of Language*. New York NY: Basic Books.
Pollard, C.J. & Sag, I.A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago IL: University of Chicago Press.
Rubino, R. & Pine, J. 1998. Subject–verb agreement in Brazilian Portuguese: What low error rates hide. *Journal of Child Language* 25: 35–60.
Saffran, J.R. 2003. Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science* 12: 110–114.
Saffran, J.R., Aslin, R.N. & Newport, E.L. 1996. Statistical learning by 8-month-old infants. *Science* 274: 1926–1928.
Swingley, D. 2005. Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology* 50: 86–132.
Tomasello, M. 2000. The item based nature of children's early syntactic development. *Trends in Cognitive Sciences* 4: 156–163.
Tomasello, M. 2003. *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge MA: Harvard University Press.
Tomasello, M., Brooks, P.J. & Stern, E. 1998. Learning to produce passive utterances through discourse. *First Language* 18: 223–237.
Velleman, S.L. & Vihman, M.M. 2002. Whole-word phonology and templates trap, bootstrap, or some of each? *Language, Speech and Hearing Services in Schools* 33: 9–23.
Vihman, M.M. 1996. *Phonological Development*. Oxford: Blackwell.
Waterfall, H. 2006. A Little Change is a Good Thing: Feature Theory, Language Acquisition and Variation Sets. Ph.D. dissertation, University of Chicago.
Waterson, N. 1971. Child phonology: A prosodic view. *Journal of Linguistics* 7: 179–211.
Wilson, S. 2003. Lexically specific constructions in the acquisition of inflection in English. *Journal of Child Language* 30: 75–115.