See Also: Domain Specificity; Genetic Basis of Language Development and Impairment; Long-Distance Dependencies; Neural Basis of Language Development; Principles-and-Parameters Framework; Specific Language Impairment (Overview); Syntactic Development: Generative Grammar Perspective; Syntax, Complex; Universal Grammar.

Further Readings

Chomsky, Noam. "Knowledge of Language as a Focus of Inquiry." In *First Language Acquisition: The Essential Readings,* B. Lust and C. Foley, eds. Malden, MA: Blackwell Publishing, 2004.

Chomsky, Noam. *Language and Problems of Knowledge. The Managua Lectures.* Cambridge, MA: MIT Press, 1988.

Chomsky, Noam. *On Nature and Language.* Cambridge, UK: Cambridge University Press, 2002.

Chomsky, Noam. *The Science of Language. Interviews with James McGilvray*. Cambridge, UK: Cambridge University Press, 2012.

Lust, Barbara. *Child Language: Acquisition and Growth*. Cambridge UK: Cambridge University Press, 2006.

Lust, Barbara. "Language Acquisition Device." In *Cambridge Encyclopedia of the Language Sciences,* P. Hogan, ed. Cambridge, UK: Cambridge University Press, 2011.

Smith, Neil. *Chomsky: Ideas and Ideals.* Cambridge, UK: Cambridge University Press, 1999.

Chunk-Based Language Acquisition

In learning to talk, children have to discover the linguistic units of their language (sounds, morphemes, words) and the ways these units can be combined to create larger patterns (inflected words and sentences). Children's progression is often characterized as a move from smaller building blocks to larger combinations: from syllables to words to multiword combinations. This characterization captures the combinatorial aspect of language learning but does not address an equally important process: the use of larger chunks to discover the units of language and the regularities governing their combination.

The idea that children use larger chunks in learning language was first formulated by Ann Peters in her seminal work on the units of language acquisition. She emphasized the role of gestalt processes in language learning and highlighted the difference between the units linguists use to analyze language and the ones children employ when learning to talk. By looking for words, linguists out on the larger, multiword sequences children extract and use in early production. The use of such units reflects the fact that infants don't hear adult speech neatly segmented into phonemes, morphemes, and words. To discover these units, children first need to break into the speech stream and identify the relevant linguistic units, a process that necessarily involves decomposing larger chunks—stretches of unsegmented speech—into smaller linguistic units. Similar whole-to-part processes play roles in children's discovery of the sounds and morphemes of their language: Children can learn about phonological contrasts by comparing whole words (whole-word phonology) and use inflected words to learn about the inflectional system in their languages.

The insight that larger chunks play roles in language acquisition was further developed in usage-based approaches to language learning—where children learn about grammar by abstracting and generalizing over stored utterances. Multiword chunks (crossing lexical word boundaries) provide children with lexically specific chunks to be used in early production and allow them to discover grammatical relations and co-occurrence patterns that hold between words. Such building blocks can be formed through undersegmentation, where a multiword sequence is first acquired as a chunk and only later properly segmented, or through chunking, where patterns of usage cause words to fuse together into one multiword unit. Both processes make the prediction that children make use of multiword chunks in the learning process.

Children's Use of Multiword Chunks

There is growing evidence that children's early building blocks include multiword chunks and that they are sensitive to the properties of multiword combinations. Children produce frozen multiword utterances at a stage where most of their other productions consist of single words. Many later productions are still not fully productive. As Elena Lieven and her colleagues demonstrate, up to 50 percent of the first 400 multiword utterances produced by 2-year-olds can be classified as frozen: Their components are not used productively but instead appear only in restricted combinations. Children's later productions are also impacted by their knowledge of larger chunks. Hence, 2- and 3-year-olds are faster and more accurate to produce higher-frequency chunks (*a drink of milk* compared to *a drink of tea*) and are impacted by chunk frequency in making syntactic generalizations. Children's morphological accuracy is also affected by the larger contexts in which words appear: 4-year-olds are more accurate at producing irregular plurals (e.g., *teeth*) inside higher-frequency chunks (*brush your teeth*). Similar patterns are seen in computational simulations. In a model that uses data-oriented parsing to parse a corpus of child speech, many of the units identified in the early stages of language production (up to age 3) were multiword ones. In another model—which uses backward transitional probabilities to identify units of language use in a corpus of child speech—children's language is better captured when the lexicon contains multiword chunks in addition to single words, reflecting the chunked nature of children's early language.

Children's use of multiword chunks is also reflected in their error patterns. For example, 4-year-olds have difficulty changing a first-person prompt such as *I think* or *I believe* into a third-person one (e.g., *he thinks*) for verbs (e.g., mental-state verbs) that predominantly appear with a first-person subject. Children's me-for-I errors (pronoun case errors such as *me do it,* where the accusative-marked pronoun is used instead of a nominative one) can be related to the proportion of preverbal uses (e.g., *let me do it*) in their input. Children are less likely to make inversion errors in questions for strings that appeared inverted frequently in the input. Over a range of constructions, children's correct and incorrect productions show sensitivity to multiword units.

Using Multiword Chunks as Building Blocks for Language Use

The reliance on multiword information is not limited to child learners. Adult speakers are also sensitive to the distributional properties of multiword chunks and draw on such information in production and comprehension. Adults have better memories for higher-frequency sequences and show reduced processing cost for object-relative clauses with more frequent subject–noun combinations. Adults are faster to recognize higher-frequency phrases compared to lower-frequency ones even when all part frequencies are controlled for (e.g., *don't have to worry* versus *don't have to wait*), suggesting that they represent frequency information about the entire complex form. Similar patterns are found in production: Speakers produce higher-frequency sequences more quickly; they are more likely to use contractions in higher-frequency sequences and show shorter phonetic duration for the same phonetic material when it appears inside higher-frequency chunks (e.g., *don't have to worry* versus *don't have to wait*). Importantly, the sensitivity to multiword frequency is not limited to idiomatic phrases or highly frequent collocations but instead is found for compositional sequences along the frequency continuum. Taken together, these findings show that multiword chunks continue to be an important part of native knowledge of language.

The Potential Differential Role of Multiword Chunks in First and Second Language Learning

One of the long-standing questions in language learning is why children seem to be better language learners than adults, despite being worse at a range of other cognitive tasks. Unlike children learning a first language, adults rarely reach native-like proficiency in a second language. However, contrary to what might be expected from a critical or sensitive period perspective, adults are not always worse than children when it comes to learning language. While adult learners clearly experience problems in many linguistic domains, they do not find all aspects of the novel language equally hard. Older learners, for instance, are generally faster and more efficient in the early stages of learning and seem to master certain domains (e.g., vocabulary) better than children. More importantly, while some facets of language are learned with relative ease (e.g., vocabulary, word order, and yes-and-no questions), other aspects—such as grammatical gender, article use, and classifiers—continue to pose difficulty even for highly proficient speakers. The currently unresolved challenge is to explain what gives rise to the specific pattern of difficulty for adult language learners.

Part of the answer to this challenge may be that adults are less likely to learn from multiword chunks and that this affects the way they learn certain grammatical relations. Whereas children are learning segmentation, meaning, and structure at the same time, adults—because of their prior knowledge and different learning situation—will learn from input that is largely segmented into words for which the semantics is already known. This tendency will affect how grammatical relations are learned by changing the information conveyed by the various linguistic elements. To give an example, when presented with the sequence *la pelota* (*the ball* in Spanish), an adult learner who already knows what a ball is can focus on the noun label at the expense of learning the pairing of the article and the noun. A child learning a first language is more likely to associate the entire article–noun sequence with the meaning of ball, thereby strengthening the link between the article and the noun.

Thus, the suggestion is that children and adults differ in their sensitivity to multiword chunks and that this can affect learning outcomes. Unlike children, the language of second language learners is often characterized as nonformulaic and is fraught with nonnative idioms and collocations. The use of collocations and formulaic expressions by second language learners is more flexible than that of native speakers, and even advanced learners produce fewer formulaic sequences than do native speakers in both spoken and written language. Indeed, when adults do acquire chunked units, they seem to use them differently in learning. Adult learners in immersion settings clearly learn some fixed expressions early on, such as greetings or requests for information. But they do not seem to use them in the same way as children, to further grammatical development. The idea that using multiword building blocks can enhance learning is supported by a recent study that manipulated the linguistic units participants were exposed to early on. In this study, adults showed better learning of grammatical gender in an artificial language when they were exposed first to sentences (multiword building blocks) and only then to individual words (single-word building blocks). They learned the association between the article and the noun better when the whole sequence was first associated with meaning, suggesting that there is an effect of early building blocks on learning outcomes.

Conclusion

Much of the work on language learning focuses on the combinatorial aspects of language: the move from smaller units to larger and more complex ones. Chunk-based learning—where larger chunks are used to discover linguistic units and the relations between them—plays an equally important role in the learning process. While understudied, such processes are found across a range of linguistic domains (phonology, morphology, and syntax). Acknowledging the importance of chunk-based language acquisition raises a new set of questions and challenges and adds new ways of accounting for how language is learned by children and adults. One major challenge lies in identifying the building blocks of language: How can one discover what children are using as building blocks? One way of addressing this question is by running computational models on child and child-directed speech to identify the most likely building blocks found in children's speech. A second challenge lies in demonstrating the role of larger chunks in learning. Multiple studies document the existence of multiword chunks in children's speech, yet there is less work showing how they impact the learning process: How do children learn to segment and analyze chunks, and how does that affect learning grammar? Here also, the combination of computational and experimental work may prove promising: Computational models can be used to identify chunks and generate predictions about their roles in learning, which can be tested experimentally. While many questions remain open, the study of chunk-based processes allows linguists to enrich their understanding of how children acquire language.

Inbal Arnon
Hebrew University
Morten H. Christiansen
Cornell University

See Also: Age of Acquisition Effects; Computational Models of Language Development; Grammatical Gender; Item-Based/Exemplar-Based Learning; Syntactic Development: Construction Grammar Perspective.

Further Readings

Abbot-Smith, K. and M. Tomasello. "Exemplar Learning and Schematization in a Usage-Based Account of Syntactic Acquisition." *Linguistic Review*, v.23 (2006).

Arnon, I. and M. Ramscar. "Granularity and the Acquisition of Grammatical Gender." *Cognition*, v.122 (2012).

Bannard, C. and D. Matthews. "Stored Word Sequences in Language Learning." *Psychological Science*, v.19 (2008).

Gobet, F., et al. "Chunking Mechanisms in Human Learning." *Trends in Cognitive Sciences*, v.5 (2001).

Lieven, E., D. Salomo, and M. Tomasello. "Two-Year-Old Children's Production of Multiword Utterances." *Cognitive Linguistics*, v.20 (2009).

Peters, A. M. *The Units of Language Acquisition*. Cambridge, UK: Cambridge University Press, 1983.

McCauley, S. M. and M. H. Christiansen. "Learning Simple Statistics for Language Comprehension and Production: The CAPPUCCINO Model." In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, L. Carlson, et al., eds. Austin, TX: Cognitive Science Society, 2011.

Wray, A. *Formulaic Language and the Lexicon*. Cambridge, UK: Cambridge University Press, 2002.

Color Cognition and Language Development

There is now extensive literature on how children acquire color words—that is, words like *red, green,* and *yellow*. This research has repeatedly shown that children seem to learn color words more slowly and with more mistakes than they learn other types of words such as nouns and verbs. By age 2, young children can readily connect object words to meanings, even from a single experience of hearing a word used in context. In contrast, even with hundreds of explicit training trials, 4-year-old children still struggle with using basic color words (such as *red, green,* and *yellow*) appropriately. Children's difficulty with color word learning is puzzling because color words are frequent in language to children and because infants have been shown to perceive colors in much the same way as adults perceive them. Children's early difficulties with color words seem so insurmountable that Charles Darwin speculated that children were initially color blind.

The course of acquiring color words is a particularly protracted one. Although children begin producing color words at around 2 years of age, they appear to use them indiscriminately and incorrectly for a period of time (ranging from months to years) and still make some color naming errors at 6 years of age. During this period, young children answer the question *What color is it?* with a color word. However, the color words they provide seem to be randomly chosen and unrelated to the property in question (e.g., they may label red, purple, and green objects as purple). And even once children begin consistently labeling certain colors correctly (e.g., always correctly labeling the color pink and never mislabeling other colors as pink), they still misidentify other common colors.

At the same time that children are experiencing difficulties in labeling colors, they may still struggle to abstract color, depending on the task. When the task demands are low, children can remember objects by their colors. In one task, children were shown a toy dinosaur and told its name was Emily. Emily was then removed, children were presented with two almost identical dinosaurs (one dinosaur the same color as Emily and one a different color), and were asked to find Emily. Even children who were not yet accurate at comprehending color words were successful in this task. Other studies with a larger number of choices show a tighter link between color words and color cognition such that children's comprehension of color words is correlated with memory for colors. Moreover, even when children can understand and use a color term, they may remain unable to selectively attend to that color without verbal cues to do so. For example, when children are shown two objects that match in color, but differ in other dimensions, and are asked to select other objects that match in the same way, children have difficulty ignoring the competing dimensions to select color matches. Altogether, these results suggest that children's understanding of color is slow to develop.

Explanations for why children are slow to learn color vary. One possibility has to do with the frequency of dimensional terms (like *color*) in the language input to young children. Early on, the input to children contains many nouns and verbs, and even though color words are frequent in speech to children, in compari¬son to nouns, the number of color words is quite low. Evidence for this idea comes from studies showing that children learn color words earlier if they attend preschool, presumably, because children receive more color word input at school. Interestingly, color words are acquired at an earlier age now than they were with previous cohorts. In 1908, IQ tests listed the normative age for naming four basic colors as 7 years of age, but by the 1970s, the normative age had shifted to 4 years of age. This suggests that difficulty with color words may stem from lack of experience and practice with color labeling, which presumably has increased for children during the last 100 years. Other explanations suggest that the difficulty can be reduced if color words appear in language structures that place the color word in more salient positions or are more transparent as to the color word's status in denoting a property of an object. For example, phrases such as "This is a red ball" indicate "red" is property of the object more clearly than phrases such as "This is red." Importantly, color words are not unique in posing difficulty for young children—children also experience difficulty with other types of abstract words and concepts (such as number, size, and space words), and not all children show protracted learning patterns.

The protracted course of color acquisition has allowed researchers to ask questions about linguistic relativity—how language influences perception