

# How Do Children Develop Syntactic Representations from What They Hear?

Elena Lieven<sup>1,2</sup>

<sup>1</sup> Max Planck Institute for Evolutionary Anthropology, Leipzig

<sup>2</sup> School of Psychological Sciences, University of Manchester

Children learn language from what they hear. In dispute is what mechanisms they bring to this task. Clearly some of these mechanisms have evolved to support the human speech capacity but this leaves a wide field of possibilities open. The question I will address in my paper is whether we need to postulate an innate *syntactic* module that has evolved to make the learning of language structure possible. I will suggest that more general human social and cognitive capacities may be all that is needed to support the learning of syntactic structure.

I start by briefly discussing precursors to language development that are developing in the first year of life: some of these are probably primate-wide skills, for instance, the capacity for distributional learning (e.g., [1]), others are probably in large part, human-specific, for instance the highly sophisticated socio-cognitive skills that one-year-olds already show (e.g., [2]).

Next, I outline an approach to language development that involves the learning of constructional schemas, both specific and abstract. Children are thought to start out with concrete pieces of language and to gradually develop more schematic constructions. All constructions are mappings between the form of the construction and a meaning, though this may not be either the full meaning or the full construction of the adult grammar. For instance a child may say *Whats that?* for months, perhaps as a request for the name of an object or perhaps as a way of getting attention without connecting the clitic *-s* to any representation of the verb, *to BE*. As their language develops children (1) learn more constructions (2) develop slots in constructions as they notice variation (3) abstract a more schematic meaning for each slot, making the constructions more abstract (4) add more slots to constructions making them more complex and (5) relate constructions to each other through distributional and analogical processes.

Many previous studies of language development have argued that children could not learn from the input because there is no surface guide to underlying constituency. In support of this, they claim that there is empirical evidence that childrens grammars are abstract from the outset. There are two major problems with assessing such claims. First, until recently, most empirical studies of language development have been conducted on very thinly sampled data. This makes it difficult to know whether either relatively infrequent utterances or the complete absence of an utterance is due to chance sampling or is really indicative of development [3]. In what follows, I report research that has largely been

conducted using dense database corpora that are orders of magnitude greater than most previous corpora.

The second problem with assessing nativist-linguistic claims is that childrens language (and that of the adults speaking to them) is often analysed in terms of the abstract categories of grammar, rather than in terms of the frequency and contexts of particular forms (morphemes, words or strings). I will demonstrate that if the utterances of children and adults are analysed at a lexically-specific level there are extremely strong relationships between the input and childrens own development and that satisfactory accounts can be given for many phenomena in language acquisition research (for instance, systematic errors) that have previously been explained in terms of pre-existing abstract syntactic knowledge.

The presentation of data falls into four parts. The first part is concerned with how children might build novel utterances from what they hear. Using a methodology that we have developed which we call *Traceback*, I assess the ways in which children could build utterances out of previously learned strings [4,5] and show how, with development, slots in constructions start to become more abstract. In this section I also present a study showing that English Child Directed Speech is very lexically restricted and that this is likely to be where childrens early lexically-based schemas come from [6]. I report briefly some current research on German and Russian CDS which, contrary to what one might expect from formal grammatical descriptions, also shows high degrees of lexical restrictiveness at the beginnings of maternal utterances [7].

However, it is important to note that while childrens language development depends crucially on the nature of the input, it is not a simple mapping from the input (for instance, the frequency of each construction) to the child's linguistic representations. This is because children build up an inventory of constructions, each a mapping of form to meaning. Childrens communicative needs and cognitive understanding play a part in this as the learning process identifies emergent forms in constructions and seeks to attach meaning to them. In the second empirical part of the paper I will demonstrate this by reference to a dense database study of the development of negation [8] and a study of auxiliary development [9].

Another reason why children are not simple frequency matchers is because constructions interconnect in ways that mean that the developing system will not be a proper subset of the adult system, but one with its own transitory states and developmental trajectory. In the third empirical section, I discuss two studies that illustrate this, one on the development of verb argument structure [10] and the other a dense database study of the development of German passive constructions [11].

In the final empirical part of the paper, I discuss data relevant to two criticisms often levelled at the usage-based approach. Firstly, it is correctly pointed out that the results come largely from English-speaking childrens acquisition and English is a very untypical language with highly restrictive syntactic word order and virtually no inflectional morphology. There are a number of reports in the literature of early and relatively error-free acquisition of morphology in morphologically-rich languages. However recent research by Aguado-Orea [12]

shows that if Spanish childrens verbal morphology is analysed at the level of specific inflections, it shows high error rates in some parts of the system and that this is closely related to the frequency of these forms in the input. A second criticism is that the dependence of the usage-based approach on frequency means that it cannot explain the sorts of systematic errors that children make with syntactically complex constructions such non-inversion and double tense marking errors in English. If children were learning the correct strings from what they hear, why would they make these errors? A recent study by Rowland [13] has shown that, in fact, these errors can be explained as a function of the frequency with which particular questions occur in the input if these are analysed lexically: highly frequent lexical strings in the input are protected from error in the childs system; errors occur when the child has less evidence as to what the correct string should be.

I conclude the paper by suggesting that the usage-based approach is by far the most promising way of making the study of language development a tractable scientific problem but that there is still a long way to go. I briefly raise some of the major challenges. These include the learning of complex morphological systems and the mechanisms underlying generalisation. Solving these problems will require considerable scientific ingenuity as well as contributions from modelling and artificial language research and the continued development of naturalistic and experimental methodologies.

In language development children build their novel utterances and their more advanced linguistic representations out of old parts [14] — here indeed is a parallel with how the evolution of language must have proceeded.

## References

1. Jusczyk, P.: The discovery of spoken language. Cambridge, MA: The MIT Press (1997)
2. Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H.: Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* **28** (2005) 675–735
3. Tomasello, M., Stahl, D.: Sampling children’s spontaneous speech: how much is enough? *Journal of Child Language* **31** (2004) 101–121
4. Lieven, E., Behrens, H., Speares, J., Tomasello, M.: Early syntactic creativity: A usage-based approach. *Journal of Child Language* **30** (2003) 333–370
5. Dabrowska, E., Lieven, E.: Towards a lexically specific grammar of childrens question constructions. *Cognitive Linguistics* **16** (2005) 437–474
6. Cameron-Faulkner, T., Lieven, E., Tomasello, M.: A construction based analysis of child directed speech. *Cognitive Science* **27** (2003) 843–873
7. Stoll, S., Abbot-Smith, K., Lieven, E.: (Lexically restricted utterances in russian, german and english child directed speech) submitted.
8. Cameron-Faulkner, T., Lieven, E., Theakston, A.: (What part of no do children not understand? a usage-based account of multiword negation) submitted.
9. Theakston, A., Lieven, E., Pine, J., Rowland, C.: The acquisition of auxiliary syntax: Be and have. *Cognitive Linguistics* **16** (2005) 247–277

10. McClure, K., Pine, J., Lieven, E.: Investigating the abstractness of children's early knowledge of argument structure. *Journal of Child Language* (2006) in press.
11. Abbot-Smith, K., Behrens, H.: How known constructions influence the acquisition of other constructions: the German passive and future constructions. *Cognitive Science* **30** (2006) in press.
12. Aguado-Orea, J.: The acquisition of morpho-syntax in Spanish: implications for current theories of development. PhD thesis, University of Nottingham (2004)
13. Rowland, C.: Explaining errors in children's questions. *Cognition* (2006) in press.
14. Bates, E., Benigni, Bretherton, Camaioni, Volterra: The emergence of symbols: Cognition and communication in infancy. New York: Academic Press (1979)