Pathbreaking verbs for developing recursion in the speech of children acquiring English

Anat Ninio
anat.ninio@huji.ac.il

The Hebrew University of Jerusalem

Abstract. The syntactic structure of sentences is represented by a tree which is built by the recursively applied operation Dependency (or Merge). Hauser, Chomsky, and Fitch (2002) claim recursion is innate and not learned. The alternative hypothesis is that children come to understand the principle of recursion on the basis of the linguistic input that transparently demonstrates the iterative nature of syntactic connection. We tested the evidence for processes of learning. Previous studies found that children first learn to use syntactic constructions with certain very frequent and prototypical pathbreaker verbs. We are now asking, are there pathbreaking verbs for recursion in children's early multiword utterances? We collected a large corpus of sentences of three words or longer possessing syntactic structure produced by 409 English-speaking young children from CHILDES (MacWhinney, 2000). We identified the word organizing each sentence, namely, the root of the syntactic tree. Most roots were verbs. There were 16 children who used only one single verb-lemma as root, and we consider these children beginners in the use of recursion. Just two verbs served for the great majority of sentences, *be* (44%) and *want* (38%). These two verbs were also used by most children who employed several different verbs as roots. Recursion has two pathbreaking verbs in child speech, *want* and *be*. Next, we analyzed three- and four-word long sentences of 508 parents from CHILDES, assuming that children may not be able to parse longer sentences and use them as their initial input for learning the principle of recursion. Close to 70% of them were built on some form of *be* as the root, but only a few (1.4%) were built on the verb *want*. The grammatical verb *be* requires two or more complements; it rarely if ever occurs without involving recursion. The results support the hypothesis that children learn the principle of recursion from the most frequent, prototypical, and generic exemplars in the linguistic input, apparently from sentences with the pathbreaking verb *be.*

Keywords: syntactic development, recursion, pathbreaking verbs, input, grammatical verbs

Introduction

This study belongs to a line of research that asks the question, how do children master syntactic structure-building? The first thing to emphasize is that although the present study represents an empirical learning approach to syntactic development, this is not a 'constructivist' or a 'usage-based' account of acquisition. We embrace a formal linguistic theoretical framework which we believe children gradually acquire, and do not hold with the idea that children store a large number of unanalyzed sentences until they submit them to statistical distributional analysis that discovers ad-hoc syntactic categories and rules. We hold that formal linguistic principles, categories and rules provide the best descriptions of people's way of constructing sentences. We believe children learn syntax from optimal exemplars which are the simplest, the most transparent, and ubiquitously encountered items of input exemplifying the relevant linguistic units and rules. Optimal exemplar learning is a general process of development found in cognitive systems as far from syntax as the immune system (Hershberg & Ninio, 2004). Potentially, such learning is one-shot, if the circumstances are optimal (see Nelson, 1987, on rare event learning). Thus, our empiricist model posits that learning is efficient and low-cost, and that our task as researchers is to find the environmental affordances for it, rather than assume that either innate principles or massive statistical analysis are involved in the process.

Our starting point is the linguistic computational system that builds sentences, the abstract core of language that consists of computational operations. We hold that this system uses a recursive combining operation called Dependency (Hudson, 1990). In the Minimalist Program a mathematically equivalent operation is called Merge (Chomsky, 1995).

The posited linguistic computational system follows the principles by which multi-element systems are constructed in computer science and mathematics. Taking syntax as the system of lawful composition of large structures from a set of combining units, it appears that a generative syntax needs just two elements: a base operation of connectivity and a process of recursion. This is a basic tenet of computer science (Abelson, Sussman, & Sussman, 1996).

The syntactic structure of sentences is represented by a dependency tree (Figure 1). The arrow leads from the head to the dependent; the head determines the character of the combination.

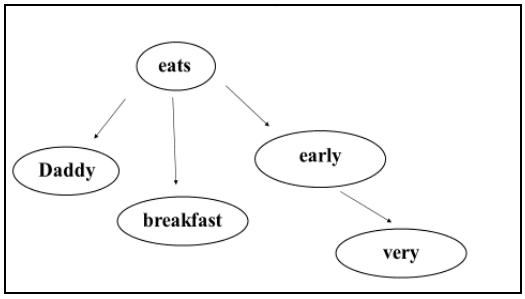

Figure 1. The syntactic structure of the sentence '*Daddy eats breakfast very early*'

Trees in mathematics are structures which are built recursively (Epp, 1995). This means the operation Dependency is applied repeatedly to connect more and more word-couples.

The base case of tree-building is a single couplet of word-nodes connected by the operation Dependency (Figure 2):

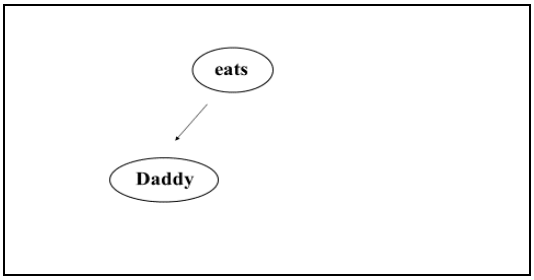

Figure 2. The syntactic structure of the combination '*Daddy eats*'

One can repeat the operation by giving one of the two existing nodes a Dependent (Figure 3). Applying the same combining operation to one of the nodes – that's the computational mechanism of recursion. Hauser, Chomsky, and Fitch (2002) claim that recursion is part of an innate Universal Grammar. The alternative hypothesis is that all syntactic principles, including recursion, are learned from the linguistic input. In previous studies, it was shown that children master the principle of dependency by learning it from transparent input sentences (Ninio, 2014a, b). Now we are asking how do children master recursion? We shall look for evidence of similar item-specific learning for recursion, from transparent input sentences. Basically any grammatical sentence three words long and longer involves the recursion of the dependency operation, by definition. When there is some reasonable doubt about the productivity of some types of multiword sentences, it is addressed explicitly.

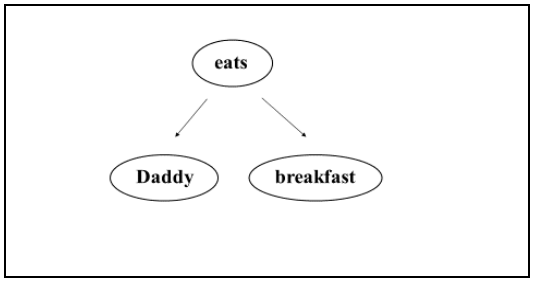

Figure 3. The syntactic structure of the sentence '*Daddy eats breakfast*'

Is recursion learned?

There are two processes of learning relevant to a novel structural principle and the actual constructions it is realized in: First, in a learning system, a new principle spreads gradually, with some first items leading the way and other items following after the while. In previous studies, it was found that children first learn to produce syntactic constructions with certain prototypical and generic verbs, which then serve as pathbreakers for other verbs (Goldberg, 2006; Ninio, 1999). Such leading verbs were found for the combinations verb-object, subject-verb-object and the like. Now we are assuming that the same phenomenon exists for the spread of novel structural *principles*. Pathbreaking verbs for verb-argument structures are not simply frequent items (as some authors believe, e.g., Theakston, Lieven, Pine, & Rowland, 2004) but, rather, they are generic exemplars that embody the relevant structural principle in a transparent way. We expect that the principle of recursion will be first expressed by children with such verb-headed sentences.

Second, in a learning system, structural principles are modeled in the input by frequent and transparent exemplars. These are usually the pathbreaking verbs children use to start the expression of the relevant principle, but even when they are not, the input affords the learning of the novel principle from highly frequent and transparent instances. In previous studies we found support for the idea that the input models novel structural principles such as the combining operation of dependency with transparent exemplars, in the shortest possible sentences (Ninio, 2014a, 2016).

We are now asking, are there pathbreaking verbs for recursion in children's early multiword utterances? Does linguistic input model transparent exemplars with high frequency? These verb-exemplars should enable, nay, force the children to acknowledge the recursive nature of the dependency operation. This means potential pathbreaking verbs should not be present in the input or in the output in sentences without recursion, *i.e.,* in two-word long sentences. This is a radical conclusion as developmental theories insist longer constructions are learned by combination or extension of children's own shorter ones (Elbers, 2000; Lieven, Salomo, & Tomasello, 2009).

Our hypotheses are:
1. Novel structural principles are first expressed by pathbreaking exemplars;
2. Structural principles are modeled in the input by frequent and transparent exemplars.

Method

Participants: children's corpus

English-language child samples were taken from the CHILDES (Child Language Data Exchange System) archive (MacWhinney, 2000), which is a public domain database for corpora on first language acquisition. All participants were observed in naturalistic, dyadic interaction of children and their parents. The observations were of normally developing young children with no diagnosed hearing or speech problems, and of their parents, native speakers of English. We restricted the contribution of each individual child to 300 spontaneous multiword sentences, starting from the first observation in which they produced multiword utterances. We collected a large corpus of sentences of three words or longer possessing syntactic structure produced by 409 children. The mean age of the children was 2;03,23 (*SD* = 4 months; range 14-42 months). We follow the tradition of researchers who examine pooled corpora of child speech for various characteristics thought to reflect on the relevant class of child speakers (Serratrice, Joseph, & Conti-Ramsden, 2003).

Results

Children's verbs at the root of multiword sentences involving recursion

In a learning system, a new principle spreads gradually, with some generic or prototypical first items leading the way. We are looking at evidence for the presence of pathbreaking verbs in children's sentences involving recursion.

We identified the word organizing each sentence, namely the root of the syntactic tree. The roots of the syntactic trees were mostly verbs. Children used 244 different verbs. The 10 most frequently used verbs were: *be, want, do, go, get, put, have, can, let,* and *see.*

Identifying beginners using a single-verb

Next, we grouped the child sample according to the number of different verbs used by the child as roots. There were 16 children who used only one single verb-lemma as root, and we consider these children beginners in the use of recursion. In this group, just two verbs served for the great majority of sentences, *be* (44% of the children) and *want* (38%). Examples 1a-j are the sentence types produced by the beginner-children with *be* as root.

1 a. He's asleep.
b. It's a kite.
c. My name is Gail.
d. That's my shoes.
e. There's the ball.
f. What is that?
g. What's this?
h. What's that?
i. Where's the ball?
j. Because they're yucky.

Examples 2a-j are the sentence types produced by the beginner-children with *want* as root.

2 a. I wanna put it back in there.
b. I want go bye bye.
c. I want leave.
d. I want snack.
e. I want that .
f. I want this.
g. I want to hold it.
h. I wanna put back in there!
i. I wanna put it back.
j. I want to put it back in there!
k. Want the big one.

Figure 4 presents the probability that children produce sentences with *be* and *want*, by the number of different roots they use, which is an estimate of the developmental phase the child is in.

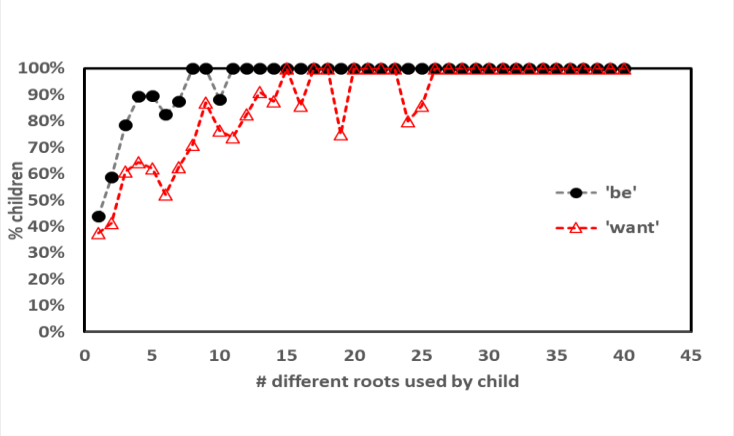

Figure 4. Probability that children produce sentences with *be* and *want,* by number of different roots they use (N=409)

The verbs *be* and *want* were also used by most children who employed two or more different verbs in their multiword sentences. The probabilities reach a plateau at 100% when the children use 10-15 different verbs as roots of longer sentences. Children use at least one of these verbs by the time they generate sentences with 3 verbs.

For comparison, the figure is repeated with the addition of the non-pathbreaking lexical verb *make* (Figure 5).

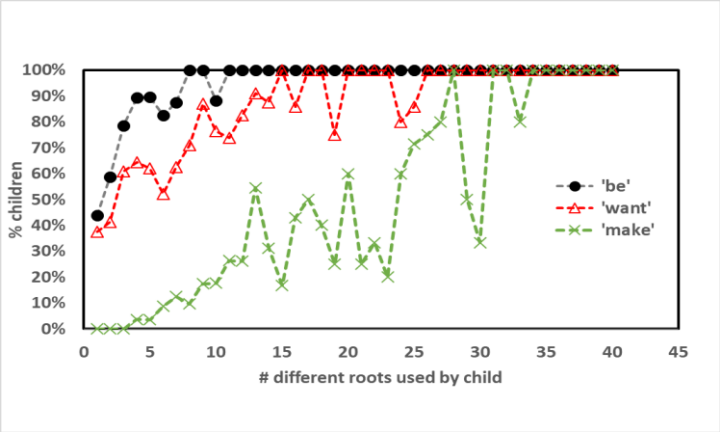

Figure 5. Probability that children produce sentences with *be, want* and *make*

A non-pathbreaking verb like *make* starts to be used as the root of sentences with recursion at a later developmental stage than *be* and *want*. Its use pattern is erratic, and reaches a plateau of 100% only when children produce 35 different verbs in such sentences. Similar patterns were observed with all other verbs serving as roots. We might conclude that recursion has two pathbreaking verbs in child speech, *want* and *be*, and other verbs.

The two pathbreaking verbs are rather different: *want* is a transitive lexical verb, with semantic arguments expressed as SVO; *be* is a grammatical verb with a subject and a second dependent which is a nonfinite verb or a predicate complement. Both dependents are usually considered not to be the verb *be*'s semantic arguments. It appears that recursion has two different prototypes, expressed in two differing syntactic constructions. Each may serve as the pathbreaker to a different subset of verbs: lexical verbs on the one side and grammatical verbs on the other.

As for syntactic valency, both verbs have a very strong requirement for two dependents, and as an indication, they rarely appear as the heads of two-word utterances where they receive only a single dependent. We checked the frequencies in children's and parents' two-word sentences, in the same samples used for the present study. Children produced 12,491 two-word sentences; among them only 1.33% was headed by *be*, and only 2.07% was headed by *want*. In the parental corpus, there were 23,204 two-word sentences; among them only 5.32% was headed by *be*, and only 0.44% was headed by *want*. The absence of such verbs from two-word long input and output sentences, demonstrates that *be* and *want* are exemplary two-dependent verbs. This characteristic turns them into prototypical roots for recursive structures.

One potential complication in considering *be* as a pathbreaking verb for recursion is that very often it is expressed in a contracted form, or cliticized on the subject. As an example we could take 1a., *he's asleep* as well as other early sentences with *be*. We should discuss the possibility sometimes raised in the literature that children learn the contracted combinations as a frozen single-word unit. In this case, the sentence including such forms does not involve recursion as it only consists of two words, for example in 1a. these are *he's* and *asleep*. If *He's* consists of a single word, there's only one dependency relation between it and the word *asleep*.

The counter-argument to this claim is that in this very developmental period, children express the same copular and temporal structures with the verb be omitted. For example, a sentence such as *He's asleep* would be expressed as *He asleep.* So called telegraphic combinations are direct evidence that children separate the subject and the copula as two words, one of which can be omitted. We checked

the incidence of telegraphic combinations by the sample and found that 309 of the 409 children of the sample, that is, 75.55%, produced telegraphic sentences with omitted copulae and auxiliary verbs. In addition, the subject is sometimes omitted as well, leaving the grammatical verb and its complement as the single expressed terms. Such sentences often alternate in adjacent turns with the fully expressed construction including the copula or auxiliary verb and its two dependents. It appears children are aware of the complex structure of cliticised and contracted forms rather that seeing them as single words. We are justified as treating *be* as a verb involving recursion when it receives a subject and a complement, whatever way it connects with the subject expression.

Parents' verbs at the root of multiword sentences involving recursion

The second part of our research plan involves exploring input sentences that embody recursion, checking if they contain frequent, prototypical, and generic exemplars from which children can learn the structural principle of recursion.

Parents' corpus

English-language parental samples were taken from the same corpora in the CHILDES archive (MacWhinney, 2000) as the child sample. There were 508 English-speaking parents in the sample. The number of utterances included from each parent was restricted to a maximum of 3,000 addressed to the child, counting from the beginning of observations. We focused on parents' three-word and four-word long sentences possessing syntactic structure, excluding vocatives or interjections. Unfinished or cut off utterances, or containing words not transcribed were also excluded.

The parents' corpus consisted of 57,820 sentences, produced by 508 English-speaking parents addressing young children. This corpus represents the 'three-and four-word long' linguistic input that young children receive when acquiring the principle of recursion. The use of pooled corpora of unrelated parents as a representation of the linguistic input is a relatively conventional move in child language research (e.g., Goodman, Dale, & Li, 2008). Table 1 presents the ten most frequent verbs used as roots for parents' three- and four-word long sentences.

Table 1. Ten most frequent verbs used as roots for parents' three- and four-words long sentences (N=57,820)

Verb	Tokens	Percent %
be	39,788	68.81
do	5,506	9.52
can	2,465	4.26
go	2,161	3.74
have	1,106	1.91
want	807	1.40
get	592	1.02
like	557	0.96
know	359	0.62
come	298	0.51

Close to 70% of the sentences were built on some form of *be* as the root, but only a few (1.4%) were built on the verb *want*. The shortest and easiest input for recursion includes a very large proportion of grammatical verbs, not lexical verbs.

Discussion

In children's speech, there are two pathbreaking verbs for producing recursive structures: *be* and *want.* Of the two pathbreaking verbs, only *be* occurs with high frequency in parental sentences three or four words long, with easily parse-able structure. Such short sentences with *be* are probably the source of the principle of recursion in general. *Want* is rare in short parental sentences; probably such sentences are probably not the input for child sentences with *want* involving recursion. Recursion in

sentences with lexical verbs as roots including *want* are probably the product of applying an already mastered principle – rather than being learned in a lexical-specific manner from the linguistic input.

References

Abelson, H., Sussman, G. J., & Sussman, J. (1996). *Structure and interpretation of computer programs* (2nd ed.). Cambridge, MA: The MIT Press/ McGraw-Hill.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Elbers, L. (2000). An output-as-input hypothesis in language acquisition. In P. Broeder & J. Murre (eds.), *Models of language acquisition* (pp. 244-271). Oxford, UK: Oxford University Press.

Epp, S. (1995). *Discrete mathematics with applications* (2nd ed.). Boston, MA: PWS Publishing Company.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford, UK: Oxford University Press.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*, 515-531.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569-1579.

Hershberg, U., & Ninio, A. (2004). Optimal exemplar learning in cognitive systems. Invited article for a special issue of *Cognitive Systems (ESSCS)* on Multidisciplinary Aspects of Learning, edited by G. Dalenoort & M. Pichat, *6*, 181-188.

Hudson, R. (1990). *English word grammar*. Oxford, UK: Basil Blackwell.

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

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum.

Nelson, K. E. (1987). Some observations from the perspective of the rare event cognitive comparison theory of language acquisition. In K. E. Nelson & A. Van Kleeck (eds), *Children's language* (Vol. 6, pp. 289-331). Hillsdale, NJ: Lawrence Erlbaum.

Ninio, A. (1999). Pathbreaking verbs in syntactic development and the question of prototypical transitivity. *Journal of Child Language, 26*, 619-653.

Ninio, A. (2014a). Learning a generative syntax from transparent syntactic atoms in the linguistic input. *Journal of Child Language, 41*, 1249-1275.

Ninio, A. (2014b). Variables and values in children's early word-combinations. *Psychology of Language and Communication, 18*(2), 106-125.

Ninio, A. (2016). Learning transitive verbs from single-word verbs in the input by young children acquiring English. *Journal of Child Language, 43*, 1103-1130.

Serratrice, L., Joseph, K. L., & Conti-Ramsden, G. (2003). The acquisition of past tense in preschool children with specific language impairment and unaffected controls: regular and irregular forms. *Linguistics, 41-42*, 321-349.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2004). Semantic generality, input frequency and the acquisition of syntax. *Journal of Child Language, 31*, 61-99.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

To cite this article: Ninio, A. (2017). Pathbreaking verbs for developing recursion in the speech of children acquiring English. In E. Babatsouli (ed.), *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2017* (pp. 213-219). ISBN: 978-618-82351-1-3. URL: http://ismbs.eu/publications-2017