**How to construct three-word-long sentences the generative way:**

**The development of recursion in young children’s syntax**

**Anat Ninio**

**The Hebrew University of Jerusalem**

Ninio, A. (2015).*How to construct three-word-long sentences the generative way:* *The development of recursion in young children’s syntax.* Full-length writeup as a complementary resource for a paper presented at a symposium on Generative Grammar at the Biennial Meeting of the Society for Research in Child Development, Philadelphia, Pennsylvania, USA, March 21, 2015. Available at <http://micro5.mscc.huji.ac.il/~msninio/three-word_recursion_longtext_online.docx>

**How to construct three-word-long sentences the generative way:**

**The development of recursion in young children’s syntax**

**Abstract**

We tested the hypothesis that children learn the principle of recursion from the linguistic input, against the alternative that children do not use this principle but merely expand their shorter sentences by adding words to them. We compared children’s first three spontaneous sentences of three words to parents’ three-word sentences and to children’s own two-word-long sentences on the words used as the syntactic *roots* structuring the sentences. Large pooled English-language corpora from the CHILDES archive were analyzed. Children produced mostly branching (tail-recursive) sentences that matched parents’ proportion of this type of recursion. We found a significant similarity in the words used as the syntactic roots structuring the sentence in children’s and parents’ three-word sentences but not children’s own two-word-long sentences. Children appear to learn a generative grammar from transparent exemplars of the linguistic input.

**Keywords**: syntactic development, recursion, generative grammar, linguistic input

Thispaper is part of an attempt to build a model of syntactic development that accepts generative syntax as its theoretical framework, but also posits that children acquire the principles of such a syntax through learning. In particular, the proffered model of development states that generative principles are learned while mastering, in an item-specific manner, input sentences constructed on those principles. The assumption is that the input contains widely available exemplars that contain transparent information on the principles responsible for syntactic connectivity; thus, cracking the code of some new syntactic principle involves interpreting the structure of individual word-combinations.

This model of development differs from most current learning theories in that it accepts formal linguistic theories of syntax, e.g., Chomsky’s (1995) Minimalist Program and Dependency Grammar (e.g. Hudson, 1990), and that it does not involve statistical processes of abstraction, extraction or distributional analysis of large amounts of the linguistic input in order to consolidate syntactic knowledge.

It follows that this model-building endeavour has two central tenets: that linguistic arguments should inform developmental theorizing regarding the units and relations we define for children’s evolving syntactic systems; and that computational considerations should be paramount in the choice of developmental processes we attribute to children. These two principles come together in our emphasis on the generativity of syntax.

A computational approach to syntactic development takes as its starting point the generativity of syntax, namely, that a finite vocabulary and a finite set of rules can generate an infinite number of different sentences. Identifying the computational processes accounting for generativity will serve to define the developmental tasks that children must accomplish in order to become competent speakers.

The insistence on generativity, namely, lack of an upper bound on sentence length and the number of different grammatical sentences in a language, is strongly associated with the work of Chomsky (1957, 1995) whose theoretical systems have always been computation-conscious. This association, however, obscures the theory-neutral character of this issue and its solution. In particular, Chomsky’s developmental approach is nativist, namely, he claims that fundamental components of human syntactic competence cannot be learned from the linguistic environment but must be passed on through the genetic system. Researchers preferring an empiricist learning theory over genetic inheritance of syntactic knowledge often shy away from the Chomskian theory altogether for this reason. This may be a counter-productive move as the computational elements of a generative grammar are theory-neutral and may serve a learning theory of syntactic development well.

In this paper, we turn to computer science to help define the fundamental terms of a generative system. 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), demonstrating that this generalization is purely formal and truly theory-neutral. Indeed, in a seminal monograph Hofstadter (1999) showed that recursive structures and processes are central in structuring all types of complex systems, so that besides syntax and computer programming, they apply also to structures in physics, biology, music and art.

At the present moment, two of the major theoretical approaches to syntax have converged on a similar if not identical definition of the base operation of syntax of natural languages. Both Chomsky's (1995) Minimalist Program and the family of theories called Dependency Grammar (e.g., Hudson, 1990) see syntactic structure as built from two-word atoms, connected by a single binary, asymmetrical combining operation -- the *Merge* or *Dependency* relation. When there are more than two words in a sentence, the syntactic structure of the complete sentence is given by the recursive application of this base operation over further words. In the basic word-couples, one word is the Head and the other is its Dependent. For example, in the sentence *Take that*, the word *take* is the head (or governor) and *that* is its dependent. (All the examples are taken from real parental and child data.) Figure 1 presents the syntactic structure of this sentence by the tree-representation customary in linguistics. Heads are connected with an arrow to their dependents.

take

that

*Take that*.

**Figure 1**. The syntactic structure of the sentence *Take that.* The head is connected with an arrow to its dependent.

Importantly, for serious linguistic reasons the merge relationship is seen as lexically-specific, defined for individual head-words, and it does not involve abstractions, categories, or indeed anything but lexical features of individual words, namely, their subcategorization (Chomsky 2000:10-11). This means that the lexical entry of the verb *take* specifies that this word gets a direct object dependent such as the word *that* in the example above. There is a need to give this information in the lexicon as there is a large degree of unpredictability whether some dependent will be expressed as a direct object (such as the object of *take*) or, possibly, as an oblique object (as in the word-combination *rely on him*). Already in the days of the Principles and Parameters model, the Chomskian theory of syntax abandoned the reference to syntactic categories and phrase-structure categorical rules once it became clear that individual words had to be marked for syntactic behavior in any case (Heny, 1979). Defining combinatory rules for form-classes and marking words in the lexicon for form-class membership created a redundancy in the system, which was declared useless and hence was removed. This move did away with abstract syntactic rules with units above the level of the lexeme, that is, phrase structure rules (see Baltin, 1989:1-3 for a discussion).

The convergence between Chomskian linguistics and Dependency Grammars in the characterization of syntactic connectivity has been noted by linguists of the two schools of thought (e.g., Epstein, 1998; Hudson, 1995). In addition, a dependency-type syntax is highly favored by computer science and in particular, Natural Language Processing, raising our confidence that this system is a strong candidate for an approach that describes the computations children learn during syntactic development.

In a generative system**,** the recursive combination of items makes it possible to create a hierarchically organized structure of any size. When applied to syntax where the base operation is that of binary merge/dependency connecting two words, recursion means that the merge/dependency operation is employed repeatedly to connect word-couples until all words participate in the connected structure of the sentence. In a structure larger than two words, one head-dependent relation is necessarily embedded in another. The smallest structure that employs recursion consists of three words; we shall define our terms using three-word sentences as our exemplars.

Borrowing terms from computer science (Abelson *et al*., 1996), we can define two different types of recursion, *true recursion* and *tail recursion.* True recursion occurs when it is impossible to complete one head-dependent combination before performing a lower one as the higher combination uses the outcome of the lower one. This is a case of required embedding, so that the overall function describing the structure of such a sentence is H(HD) – the higher head gets the lower head-dependent couplet as its dependent. Figure 2 presents the tree for the sentence *Look at me!* that possesses such a structure.

look

at

me

*Look at me!*

**Figure 2.** The syntactic structure of the sentence *Look at me!* Heads are connected with an arrow to their dependents.

In an informal description we may say that the verb *look* gets the prepositional object *at me* as its dependent. Until we compute the meaning of the whole phrase, we do not know where it is that the speaker wants the hearer to look. More formally, such a function builds up a deferred operation, so that the dependent is not defined until the lower head-dependent couplet is computed. The computation cannot be performed in any other order, as the dependent is the phrase defined by the subtree (*at me*), not the single word (*at*) which is the phrase’s formal head. Another example may be the sentence *Want this book* in which the verb *want* gets the determiner phrase *this book* as its dependent. Again, the lower head-dependent couplet *this book* needs to be computed before the verb can get its rightful dependent.

In syntactic trees such as those in Figures 2 and 3, all words are connected, and they all have a head -- except for the highest word that only has dependents. That is the *root* of the sentence, the word all others are subordinate to. In Figure 2 the root is *look*, in Figure 3 it is *sits*. This is the word responsible for structuring the sentence. Although roots of complete sentences are always verbs, elliptical sentences such as answers to questions can have other roots such as determiners and prepositions. For example, the answer to *Where is the bird?* may be *On the roof*. Some root-words such as verbs can have many dependents, making branching or tail-recursion possible, others such as a preposition may not. Thus the identity of the root may have an important influence on the type of recursion observed in the tree.

Tail recursion of three words (aka *branching*) consists of two head-dependent relations branching out of the same head-word. Some examples are *Bunny is sleeping* and *I got you!* Figure 3 presents as an example the syntactic structure of the sentence *Daddy sits here*.

sits

Daddy here

*Daddy sits here*.

**Figure 3.** The syntactic structure of the sentence *Daddy sits here.* Heads are connected with an arrow to their dependents.

The first fundamental principle of a generative syntactic system that should be mastered by children is merge/dependency as a base operation connecting two words in an asymmetrical binary relationship. Two-word long sentences are children's earliest productions of syntactically connected word-combinations and they comprise the great majority of their productions for a considerable time.

As two-word couplets constitute atomic units of syntactic structure, it means that to learn to construct them, young children need to master one single basic production strategy: to produce in one turn of speech two words that are in an asymmetrical relation, with one word serving as the dependent of the other. We are following previous publications such as MacWhinney (1982), Ninio (2006), Powers (2002), and Van Langendonck (1987) and propose that children’s syntactic development is best conceptualized as the gradual mastery of the principles and specific details of a generative syntax, right from the start of two-word combinations.

Ninio (2014a) examined English-speaking parents’ two-word utterances containing a verb and its subject, object or indirect object, that is, the core grammatical relations. The results showed that such sentences contain transparent information on the binary dependency/merge relation responsible for syntactic connectivity as the verbs are inevitably coupled with pronouns and proper names which cannot be anything but referential expressions specifying a dependent of the verb. In addition, pronouns and proper names are natural variables for dependents as they are indexical expressions, receiving a different value in each different context, making generalization to other contexts an immediate possibility. In a second study, a large sample of children were found to use the same verbs as the parents in the great majority of their early sentences which express those same core grammatical relations, suggesting that children learned the basic principle of merge/dependency needed to produce two-word sentences, as well as the specific two-word constructions, from parental two-word sentences. Interestingly, the terms used by the children as subjects and objects did not match the distribution of the ones in the parental input, supporting the conclusion that children learned that certain verbs take, for example, an object, but that they were free to specify the object according to current circumstances (Ninio, 2014b).

After learning to combine two words by the merge/dependency operation, the next developmental task children face is to learn to apply this operation recursively in order to build syntactically connected three-word sentences, and, with repeated recursion, sentences of any length (Hauser, Chomsky, & Fitch 2002). In the present study, we are hypothesizing that children master the principles and details of recursive structure-building by a route similar to the way they master the basics of the system for the construction of two-word sentences – that is, by learning them from the appropriate parental input in an item-specific manner.

According to our generative hypothesis, then, children’s first three-word sentences represent a serious phase shift in their syntactic system. As the new principle required for connecting three words – recursion – does not exist in two-word sentences, children turned to parental three-word or longer sentences when mastering the new concept and the new constructions it is used for. This conceptualization tags children’s first three-word sentences as radical innovations rather than simple extensions of their two-word sentences. In the absence of such a presupposition, an alternative hypothesis can be offered, namely, that there is nothing revolutionary in three-word sentences, and they can probably be easily derived by children from their two-word sentences by extending them with the simple addition of another word. Such claims are indeed found in the literature where children’s sentences are said to be mostly derived from their own previous sentences, by some minimal change such as the addition of another word (Lieven, Salomo, & Tomasello 2009). These authors’ work suggests that once children have started to combine words into 2-word utterances, the principles needed to produce 3-word utterances are already in place, and therefore the one can serve as the basis for the other. This divergence in the perceived role of parental input and of children’s shorter sentences in the buildup of children’s initial three-word sentences makes it possible to define an alternative hypothesis to our learning model. Under our generative hypothesis, children’s initial three-word sentences should be more similar to parents’ three-word sentences than to children’s two-word sentences. According to the alternative hypothesis, children’s initial three-word sentences should be more similar to their two-word sentences than to parents’ three-word sentences.

According to an analysis of the learning task, the optimal input for learning recursive constructions should be parents’ three-word long sentences. These are long enough so that they demonstrate recursion – a pattern that cannot be observed in two-word sentences -- and short enough so that children do not have to decipher the structure of sentences several steps beyond their current level of knowledge such as four-word or longer strings. We are therefore hypothesizing that parents’ three-word sentences provide highly learnable input for discovering how to construct three-word sentences.

In order to test the learning hypothesis against the alternative hypothesis, we will compare children’s first three spontaneous 3-word long sentences to parents’ three-word sentences in terms of the distribution of recursive constructions. We hypothesize that as they start to produce three-word sentences, children will use the two types of recursion (tail and true recursion) in proportion to their distribution in the parental input of three-word sentences. In addition, as we assume children learn lexically-specific concrete combinatory patterns and not abstract principles, we hypothesize that children will be similar to parents in the details of their initial three-word constructions, as will be demonstrated by a similar distribution of the root words heading the three-word constructions they produce (for a recent review of the input frequency effect see Ambridge, Kidd, Rowland, & Theakston 2015) . We shall also compare children’s initial three-word sentences to their two-word sentences and hypothesize that they will be less similar to these than to parents' 3-word sentences.

To test these hypotheses, we collected a large corpus of children's first three three-word sentences with syntactic connectivity observed in the recorded observations, and compared them to children's two-word sentences and to parents' three-word sentences.

**Method**

*Participants*

English-language parental and 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. This resulted in the selection of parents and children from 33 research projects in the CHILDES archive: the British projects Belfast, Howe, Korman, Manchester, and Wells, and the American projects Bates, Bernstein-Ratner, Bliss, Bloom, 1970 and 1973, Brent, Brown, Clark, Cornell, Demetras, Feldman, Gleason, Harvard Home-School, Higginson, Kuczaj, MacWhinney, McMillan, Morisset, New England, Peters-Wilson, Post, Rollins, Sachs, Suppes, Tardif, Valian, Van Houten, and Warren-Leubecker (MacWhinney, 2000). From these projects, we selected the observational studies of 471 different parent-child dyads. In 35 of the studies there were two active parents interacting with the target child, resulting in a parental sample of 506 different parents.

In order to avoid severely unequal contributions to the pooled corpus, the number of utterances included from each parent was restricted to a maximum of 3,000, counting from the beginning of observations. We have excluded the speech of parents addressed to other adults present in the observational session or on the telephone, as this speech may be ignored by young children because of unfamiliar subjects. Contextual comments were checked in order to ascertain that we included only spontaneous utterances from target parent to target child. The resultant parental corpus contains almost 1.5 million (1,470,811) running words.

We focused on parents’ three-word sentences possessing syntactic structure, excluding utterances where one of the three words was a vocative or an interjection. Unfinished or cut off utterances, or containing words not transcribed were also excluded. The parents’ corpus consisted of 38,512 three-word sentences, produced by 506 English-speaking parents addressing young children. This corpus represents the three-word long linguistic input that young children receive when acquiring syntax. 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). Multiple speakers of child-directed speech may provide a good estimate of the total linguistic input to which children are exposed, which includes, besides the speech of the individual mother or father, also the speech of grandparents, aunts and uncles, older siblings and other family members, neighbours, care professionals, and so forth, represented in our corpus by the speech of mothers and fathers unrelated to the individual child. The pooled database represents the language behaviour exhibited by the community as a whole when addressing young children.

*Children's sample***.** Samples of two- and three-word utterances from 471 children were taken from the same English transcripts in the CHILDES archive from which we took the parents’ speech. The mean age of the children was 2;01,03 (*SD* = 4 months; range 14-42 months). We restricted the contribution of each individual child to 300 multiword sentences, starting from the first observation in which they produced multiword utterances. Children's utterances were included only if they were spontaneous, namely, not immediate imitations of preceding adult utterances. For each utterance marked in the original transcriptions as one uttered by the child, we checked the context to make certain that the line was indeed child speech (and not, for example, an action description or parental sentence erroneously marked as child speech). The size of the resulting pooled corpus is 194,359 running words. The children’s corpus contained 101,064 utterances; this makes the mean MLU in words 1.92. Similar to the group of parents, we are treating young children acquiring English as their first language as a homogeneous group, as far as the important characteristics of their syntax is concerned. In this, 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 (Radford, 1990 ; Serratrice *et al.*, 2003).

For this study, we selected the children who were in most probability at the at the very start of three-word speech, as the whole sample consisted of children who produced as yet very few three-word utterances. Of the total sample of 471 children, 350 children produced at least three sentences of three words in syntactic combination, that is, excluding vocatives, interjections, or syntactically unrelated words, during the period of observation sampled in the study. The other children produced none or two at the most. This generated a total corpus of 1,050 three-word sentences in syntactic combination said at the start of three-word speech, each child contributing just three sentences. The first three sentences of this kind found in the recorded observations are unbiased estimates of the kind of sentences that start the production of three-word syntactic speech. The hope is that they represent faithfully the first set, although of course it cannot be established that they were absolutely the first sentences said.

Focusing on the first three sentences observed for each child but no later ones was done to make the sample represent the entry point to three-word speech. The goal of the study was to understand the transition period into three-word speech that would reflect the acquisition of the syntactic principle necessary for producing such sentences. The first set of three-word sentences could reveal the type of constructions that serve as the functional source for learning the principle, probably because they reflect the structural principle in a particularly transparent way. Hopefully, this will illuminate the process of acquisition.

We also analyzed the same children’s two word-long sentences possessing syntactic structure. The children produced a total of 11,642 two-word sentences (mean 33.26, *SD* = 21.86).

*Data analysis*

*Syntactic annotation for grammatical relations***.** Sentences were parsed manually for syntactic structure. We based our dependency analyses on the detailed descriptions of Hudson’s English Word Grammar (Hudson, 1990) with its online update (Hudson, 2014). We also consulted descriptive grammars of English, and in particular Quirk, Greenbaum, Leech, and Svartvik (1985).

For each sentence, the root (namely, the highest element syntactically) of the dependency structure was identified and subsequently tagged for form class membership (see below). In three-word sentences, we classified the type of recursion employed for structuring the sentence. When the root received two dependents, we classified the recursion as tail recursion; when the root had a single dependent and its dependent received a dependent of its own, we classified the recursion as true recursion. For examples, see Figures 1 and 2.

Syntactic annotation of the sentence was done by graduate students at the Hebrew University with training in linguistics. It relied on extensive coding instructions and a very large collection of annotated exemplars. We checked for reliability by having three pairs of coders blindly recode 1,900 utterances produced by four different parents and two children. A checking of all reliability codes showed that the agreement of each coder with the others was above 95%, based on codes actually given by the relevant pairs of coders. Throughout coding, all problem cases were discussed and resolved. Ultimately, each coded utterance was double-checked by another coder.

*Classifying roots for form-class.*The root-words were classified according to categories of form-classes. Table 1 presents the form-classes employed.

**Table 1.** Form-classes used to classify roots of sentences.

**---------------------------------------------------------------------------------------------------------------**

|  |  |  |
| --- | --- | --- |
| Form-class symbol | Definition of form-class | Examples of words in category |
| Aux-V | Auxiliary verbs including copula and dummy verb | be, is, was, can, may, have (auxiliary), do |
| Lex-V | Lexical verbs | come, go, want, get, see, eat |
| CN | Common nouns | baby, ball, bottle, bear, bird |
| PN | Proper nouns | Anna, David, Mommy, Nina |
| PR | Pronouns, e.g. demonstratives, indefinites, interrogatives | I, he, they, this, that, somebody, something, one, who, which |
| DT | Determiners, e.g. articles, numerals, possessive pronouns, possessors | a, an, the, this, that, some, two, my, your, John’s, no |
| PP | Prepositions | to, from, in, on, like, for, by |
| AJ | Adjectives | big, little, red, wet, good, bad |
| AV | Adverbs | very, here, there, now |
| TO | TO-infinitive | to |

------------------------------------------------------------------------------------------------------------ Form-class tagging was done automatically, employing a list of tagged words serving as roots of sentences which was prepared manually and subsequently run on the lines of the corpus. The accuracy of the tagging was checked manually against the linguistic context, and errors were corrected.

**Results and discussion**

*Children are following parental models*

To establish that children indeed follow parental models of type of recursion, we compared the type of recursion employed to structure children’s first three three-word sentences to parents’ three-word sentences. Table 2 presents the results.

**Table 2**. Children’s first three three-word sentences and parents’ three-word sentences by type of recursion.

Speaker Total Tail recursion True recursion

Tokens Tokens % Tokens %

Children 1,050 799 76.10 251 23.90

Parents 38,512 32,809 85.19 5,703 14.81

The distributions are quite similar; in both corpora the great majority of sentences are structured with tail-recursion, with a minority using real recursion. We might summarize that the comparison does not refute the hypothesis that children follow parental models of recursion.

We compared the distribution of the syntactic root-word by form-class in children’s three-word sentences to parents’ three-word sentences and to children’s two-word long sentences. Table 3 presents the results.

**Table 3**. Distribution of the syntactic root-word by form-class, in children’s and parents’ three-word long sentences, and in children’s two-word long sentences.

**Form-class Children 3-words Parents 3-words Children 2-words**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**symbol Frequency % Frequency % Frequency %**

**---------------------------------------------------------------------------------------------------------**

Lex-V 547 52.10 14,362 37.29 3,814 30.54

Aux-V 407 38.76 21,044 54.64 337 2.70

DT 43 4.10 1,467 3.81 4,091 32.75

PP 35 3.33 1,167 3.03 1,400 11.21

CN 8 0.76 204 0.53 1,887 15.11

AJ 6 0.57 119 0.31 366 2.93

AV 3 0.29 43 0.11 101 0.81

PR 1 0.10 66 0.17 330 2.64

PN 0 0.00 25 0.06 155 1.24

TO 0 0.00 15 0.04 9 0.07

----------------------------------------------------------------------------------------------------------

Total 1,050 38,512 12,490

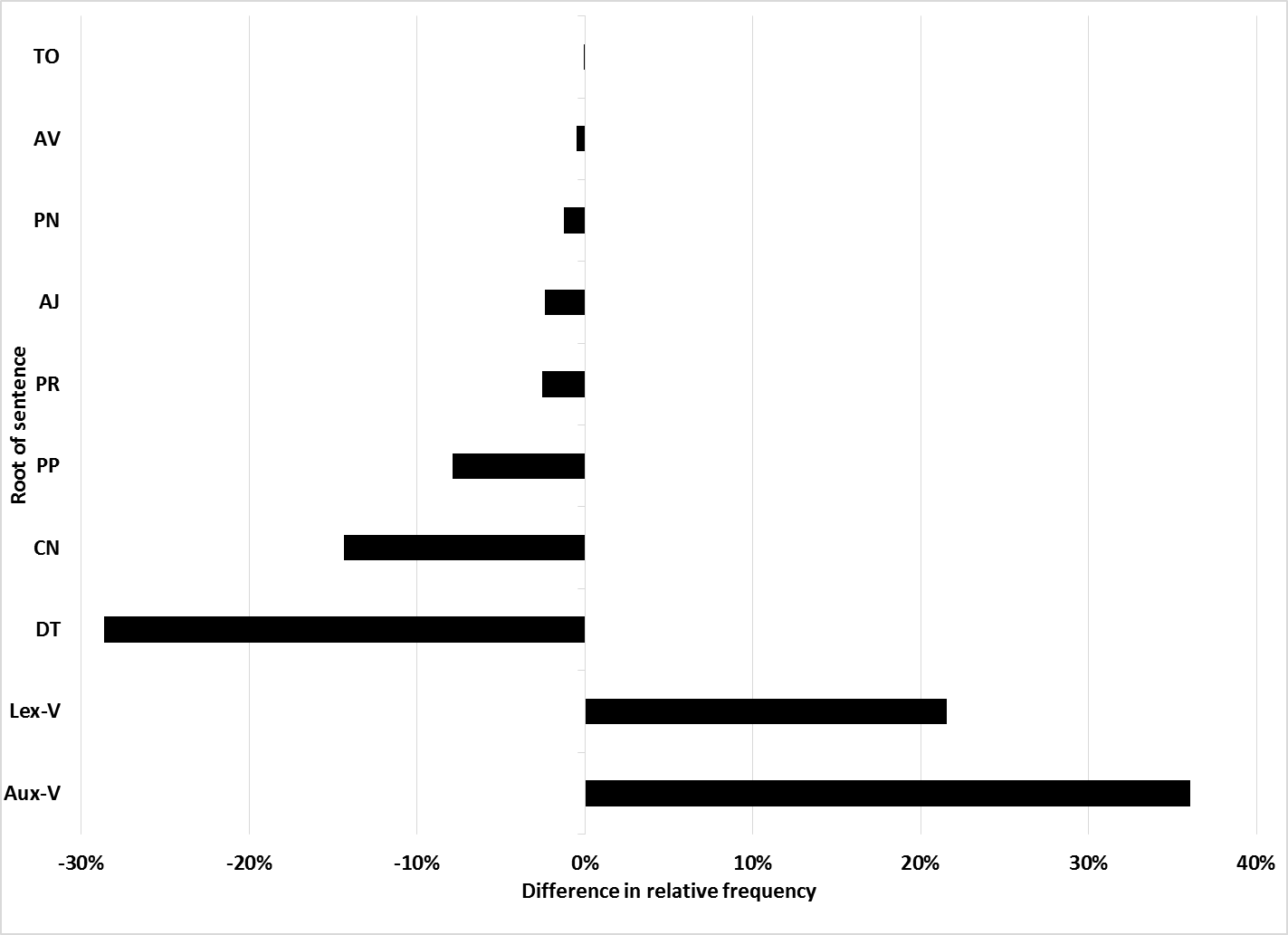
----------------------------------------------------------------------------------------------------------

*Note*: Lex-V = lexical verb; Aux-V = auxiliary verb; DT = determiner; PP = preposition; CN = common noun; AJ = adjective; AV = adverb; PR = pronoun; PN = proper noun; TO = *to* infinitive marker

We computed Pearson correlation coefficients to test the hypothesis that children’s first three three-word sentences are more similar in their syntactic roots to parents’ three-word sentences than to children’s two-word sentence. We found that there was a large difference between the two correlation coefficients. Roots of three-word sentences by parents predicted children’s roots in their earliest three-word sentences with a highly significant correlation (*r*(9)=0.939, *p*<0.001), accounting for 88.10% of all variance, whereas the roots of children’s two-word sentences did not predict them significantly (*r*(9)=0.415, *p*>0.05), accounting for a mere 17.26% of variance. These results support the hypothesis that children learn to produce three-word sentences from parents’ three-word sentences in an item-specific manner, using the sentence-types modelled in the input, rather than extending their own two-word sentences by adding another word to them. Had they produced their three-word sentences by using some abstract production rule referring to any root-word getting recursive dependents rather than rules learned in an item-specific manner for particular words, there would not have been a correlation of the roots they used with parents’ roots, as that very high correlation can only be the result of a frequency-effect on the probability of children learning a particular word with its two syntactic subordinates.

This result is similar to the one obtained in a study regarding children’s two-word sentences expressing core grammatical relations, where the distribution of the verbs heading the sentences was similar to that of parents’ same-length sentences (Ninio, 2014a). As the child sentences in these studies are spontaneous and productive, children need to know how to generate them using the syntactic principles of dependency and recursion. Thus, the similarity is most likely to be the result of learning from the same-length input, not of merely imitatively matching their sentences to same-length parental ones.

We next examined the mismatch between children’s three-word sentences and their two-word sentences to pinpoint the reason for the low correlation between them. Figure 4 presents the differences in relative frequency of the form-class into which the sentences’ roots fell. The graph presents the change in the relative frequency of each form-class from children’s two-word sentences to children’s three-word sentences.



**Figure 4.** Change from children’s two-word sentences (N=12,490) to their three-word sentences (N=1,050) in the relative frequency of sentence roots belonging to different form-classes. TO = *to* infinitive marker; PN = proper noun; PR = pronoun; AV = adverb; AJ = adjective; PP = preposition; CN = common noun; DT = determiner; Lex-V = lexical verb; Aux-V = auxiliary verb.

It appears that children build their three-word sentences around verbs much more than their two-word sentences, whereas non-verbal root-words are much less frequent in their three-word sentences than in two-word ones. To see this trend more clearly in all three corpora, we collapsed verbal roots into one more general category and non-verbal ones into a second. Table 4 and Figure 5 presents the relative frequency of verbs and non-verbs that serve as syntactic roots of children’s first 3-word sentences compared to their 2-word sentences and to parents’ 3-word sentences. Figure 5 is fully redundant with Table 4 and it is presented merely because it is easier to process visual information.

**Table 4. F**requency distributions of the highest element of the sentence, its root, in the three corpora, by verb status.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Root of sentence | Children's 2-word sentences | | Children's 1st 3 3-word sentences | | Parents' 3-word sentences | |
|  | Frequency | % | Frequency | % | Frequency | % |
| Verb | 4,151 | 33.23 | 954 | 90.86 | 35,406 | 91.93 |
| Not verb | 8,339 | 66.77 | 96 | 9.14 | 3,106 | 8.07 |
| Total | 12,490 |  | 1,050 |  | 38,512 |  |



**Figure 5.** Relative frequency of verbs and non-verbs serving as syntactic roots of children’s first 3-word sentences (N=1,050) compared to their 2-word sentences (N=12,490) and to parents’ 3-word sentences (N=38,512).

Children’s earliest three-word sentences match parents’ three-word sentences in proportion of verbs as roots (91% and 92%), and both are considerably different from children’s two-word sentences (33%).

To understand why children move from a majority of nonverbal roots in their two-word sentences to a majority of verbs in their three-word sentences, we hypothesized that verbs enable the easier kind of recursion, namely, the branching tail-recursion, whereas non-verb roots are more likely to involve the more difficult true-recursion that involves embedding. To check this possibility, we broke down tail-recursive and true-recursive sentences in parents’ three-word speech by type of root-word, distinguishing between verbal and non-verbal roots. Table 5 presents the results.

**Table 5**. Sentences with verb and non-verb roots by type of recursion in parents' three-word sentences.

Root /Recursion Verb Not verb

Frequency % Frequency %

Tail-recursive 32,089 90.64 720 23.15

True-recursive 3,313 9.36 2,390 76.85

Total 35,402 3,110

It appears that sentences headed by verbs in parental speech are almost exclusively tail-recursive or branching, e.g., *I can see*, whereas sentences headed by non-verb roots in the majority of cases involve true-recursive structuring e.g., *A good girl.*

As we hypothesized, when children switch to verbal roots for their initial three-word sentences, they make use of the affordance to generate the easier branching structure with a verb as the organizer of the sentence. Table 6 presents the distribution of children’s three-word sentences with verb and non-verb roots by type of recursion.

**Table 6**. Sentences with verb and non-verb roots by type of recursion in children's three-word sentences.

**Root /Recursion Verb Not verb**

**Frequency % Frequency %**

Tail-recursive 788 82.60 11 11.46

True-recursive 166 17.40 85 88.55

Total 954 96

We tested the relation between root types (verbs and non-verbs) and types of recursion in children’s speech by a chi square test. There was a highly significant relationship between root type and recursion type, X2 = 248.64, *p*<0.001. Verb-rooted sentences are mostly structured by tail-recursion, non-verb rooted sentences mostly employ true-recursion. This also means that children’s three-word sentences very seldom contained their two-word verb-less sentences as components, as the branching structure cannot accommodate a two-word determiner phrase or prepositional phrase.

**Input sentences provide transparent models of syntactic principles**

According to our model of syntactic development, in order to support the learning of structural principles from exemplars, three-word sentences in parental input should provide transparent information about the syntactic principle of recursion that had structured the relevant sentences. It appears that transparency of structure is achieved in the three-word input by the use of verbs as roots of sentences, and, in particular, by verbs possessing two semantic arguments mapped to syntactic dependents. As we saw in Tables 3 and 4, 92% of parents’ sentences are built on verbal roots, including lexical verbs, auxiliary verbs and copulas. Table 5 showed that most verbs (91%) head a sentence with a branching tail-recursive structure. Checking the most frequent verbs used by parents, it appears that these require two dependents on semantic grounds. The auxiliaries *be, can* and *have* generate sentences such as *Bunny is sleeping; Daddy is coming; are you hiding?; I can see/ move/ tell; has she finished; has it gone?; have you forgotten/ pooped?; what has happened?.* The auxiliary and dummy verb *do* is at the root of sentences such as *I did not; I did see; I did so; I do too; did you burp?* *do you know?* The copula *be* heads sentences such as *What's it/ that/ this?; it is bad/ beautiful/ big/ blue/ cold/ cute/ dirty/ empty/ good/ green* and so forth. These grammatical verbs need a subject and a predicate complement, and their lexical semantics suggest as much, so that, for example, the meaning of the copula *be* explicitly signifies that there is an identity between its two dependents.

Other frequent roots of branching sentences are the transitive verbs *get, want, do, see* and *say* that generate sentences such as *They got some; Mummy got hers; I got it!; I got you!; I get it; I want cookies; I want dinner; Baba wants it; Cathy wants you!; Daddy did it; you do that; you do it; I saw you; I see her; I see shoes; cow says moo; he says that?; I said no.*

These verbs, too, possess lexical semantics that translates to a need for a subject and an object. Lastly, among the frequent roots of branching sentences in the input thereare themotion verbs *come* and *go* that generate sentences such as *This comes off; who came out?; you come here; Bob went away; Daddy went upstairs; down it goes.* These are intransitive verbs that nevertheless get two lexically expected dependents, based on their semantic/syntactic need for a subject and, optionally, also a directional locative, or else they get particles with an adverbial meaning. It seems that in all these cases, the branching structure is made transparent by the verbs’ lexically specified semantic relation with the two dependents.

In many cases, roots also help the child to figure out he structure of real-recursive sentences. As we saw in Figure 2, verbs and adjectives that receive oblique objects as their complements needs to be supplied with the whole prepositional phrase in order to be grammatical. The most frequent word serving as the root of a real-recursive sentence was in fact the verb *look*; in the majority of sentences, *look* received an oblique object consisting of the preposition *at* followed by some nominal. Some examples are *Look at Mommy, look at it, look at me,* and *look at that.* Examples of other verbs and adjectives getting oblique objects are *Smile at Mumma; mad at me?; keep at it; wait for it; look for it; good for you!; ready for picture?; time for diaper; think about it* and *Sorry about that.*  A further pattern of verb governing an embedded phrase is found in the parental input in sentences where the verb receives the infinitive marker *to*, which in turn requires a verbal complement of its own. Examples are *Care to dance?; got to hurry; likes to play; need to go; need to wait; trying to walk?; want to come?; want to cook?; want to try ?.* Just like in the case of the branching dependents of our prototypical tail-recursive sentences such as *I see you*, receiving such an “embedding dependent” is a lexical feature of the relevant verbs and adjectives. It is possible that such exemplars would provide the child with an insight into the principle of real recursion, helped by the idiom-like character of combinations with oblique objects. There are of course other kinds of real-recursive sentences in the input as well, such as determiner phrases like *A good girl* and verbs with extended direct objects such as *Want this book*? and *See the lion*. It is possible that children figure out their structure by identifying the lower syntactic couple as well as the lexical valency of the root word. According to our data, however, such complex exemplars were a small minority both in parental three-word speech and in children’s initial three-word sentences. It is possible that for the explanation of how children learn to generate such patterns, we need to look at longer sentences in the input.

**General discussion**

Our results support the hypothesis that children learn their earliest three-word combinations mostly on the basis of parental three-word combinations, rather than derive them by combining or expanding their own two-word sentences. The very high correlation with parents’ three-word sentences demonstrates a strong frequency effect that parental input exerts on children’s earliest three-word long sentences. By contrast, the low correlation with children’s two-word sentences shows that the distribution of these shorter sentences does not predict what children will say in their earliest three-word sentences. Our interpretation is that in order to build more than a single head-dependent couplet, children have to learn new principles, and these are demonstrated in parental three-word sentences and not in two-word long sentences. Three-word sentences in the input are mostly headed by verbs; children’s two word sentences are mostly headed by words other than verbs. Parents do not on the whole demonstrate how to build three-word structures headed by determiners, prepositions or common nouns. The input thus does not model for children how to construct more complex non-sentential phrases using three words using the same roots as two-word long phrases. Instead, the relevant parental speech models how to build three-word sentences headed by verbs. Apparently, it is from these that children learn how to go beyond the single head-dependent couple expressed in a two-word utterance.

The results of this study support a model of syntactic development according to which children acquire a generative syntax with basic principles similar to those of Dependency Grammar and the Minimalist Program: a basic binary combinatory operation along the lines of merge/dependency for generating two-word long syntactic atoms, and recursion for generating sentences longer than two words.

We found that there is a way for children to learn the abstract principles of syntactic connectivity from the linguistic input, with a computationally simple and non-demanding strategy. Children can derive syntactic principles from the shortest possible input sentences that demonstrate the relevant principle: two word sentences for deriving the asymmetrical head-dependency base operation, and three word sentences for deriving recursion. Two word sentences with a verb and one of its core arguments are transparent regarding the asymmetry of relationship, as the vast majority of the referential expressions used in such short sentences are pronouns and other indexical expressions that cannot serve other goals but to point to some entity, whereas the verb of the sentence possesses a complex semantics like a mathematical function on arguments, that requires the specification of its semantic arguments by other words in the sentence. The juxtaposition of pairs of words with these clearly demarcated complementary features defines the two-word sentence with core syntax as being in an asymmetrical relation where the verb is the head of the referential expression that is its dependent (Ninio, 2014a). Three word sentences are transparent regarding the principle of recursion because in most input sentences the root is a verb with a valency requirement for two syntactic dependents, so that the base operation of merge/dependency is repeated either for the head (generating a tail-recursive structure) or for the dependent (generating true recursion). The abstract principles of syntax are embodied in these simple structures, and the syntactic code can be cracked on a concrete and interpretable level.

From a computational point of view, such a learning process – making use of the optimal input available in single sentences for the child’s zone of proximal development – has a considerable advantage over much more cumbersome processes offered in models of statistical learning. Using single transparent sentences as input to development, children do not need to perform two computational acts that would place considerable load on their cognitive system: collecting and storing in memory a large number of unanalyzed raw sentences, and putting this large data set under a distributional analysis in order to extract from it the units and categories of a productive syntax (e.g., Bannard, Lieven, & Tomasello, 2009). This broad generalization ignores recent statistical learning approaches that emphasize now-or-never processing (Christiansen & Chater, in press) or extraction and integration (Erickson & Thiessen 2015) rather than statistical analysis over stored exemplars in memory.  These contemporary accounts of statistical learning may be better compatible with the item-based learning mechanisms proposed in the current model. Putting aside the question whether such a computational effort would generate a syntax that could compare to the formal syntactic theories of linguistics, it is questionable that children can be expected to be able to perform the computations expected of them under such models of development. Paradoxically, a theory of development based on formal linguistics is able to offer a learning process that possesses a higher likelihood of psychological reality than a purely statistical model.

Apart from the a priori advantages of a model of acquisition grounded in formal linguistics, the present study provides some empirical evidence supporting it. The contrasting correlations of children’s sentence types with parents’ and with their own two-word speech showed that the prediction derived from formal syntactic theory accounts much better for the developmental data than an alternative that does not acknowledge the theory-derived sharp innovation in the structure of three-word sentences, namely, recursion. A statistical-distributional oriented theory of acquisition cannot see the phase-shift in three-word sentences that necessitates turning, again, to the linguistic input for novel principles that a child does not yet possess in the so-called two word stage, namely, when producing only two-word long sentences.

In addition, we made use of a distinction between two sub-types of recursion in syntax based on a formal computational analysis employed in computer science, true recursion and tail recursion. We found that the distribution of the two types of recursion in children’s initial three-word sentences is very similar to that of parents’ three-word sentences, so that the majority of sentences in both corpora are structured with tail recursion (aka branching) and the minority, with true recursion. The very high proportion of tail recursion in parental speech may indicate that this pattern is typical of three-word long sentences, or that, as we suspected, it is simpler to generate than true recursion. In any case, the pattern of results indicates that the theoretical distinction derived from a mathematical system of computation is meaningful for both usage and acquisition.

The concept of recursion only makes sense when there is a base operation that can recur and generate all length of sentences, and in the linguistically-oriented model of development, that is the binary merge/dependency operation. Our results strongly support the attribution to children of a generative syntax with a base operation and recursion. The results make it possible for us to adopt a linguistically grounded characterization of children’s early syntactic system, employing as its basic syntactic combinatory operation the asymmetrical head-dependent relation of merge (or dependency), and continuing the evolution of the syntactic system by learning to use recursively the basic combinatory operation in the production of syntactically structured three-word and longer sentences. In this study, we offered a possible route by which children may learn the principles of a generative syntax in a computationally easy way. Thus, it appears possible that children learn a true version of adult syntax as described by the generative linguists, both Chomsky’s (1995) Minimalist Program and Dependency Grammar (Hudson, 1990).

**Acknowledgments**

Portions of the research reported here are to be presented at the Biennial Meeting of the Society for Research in Child Development, Philadelphia, Pennsylvania, USA, March, 2015. Thanks are due to the parents and children who participated in the observational studies, and to the researchers who contributed the corpora to the CHILDES archive.

**Funding**

This work was supported by the Spencer Foundation [grant number 200900206].

**References**

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

Ambridge, A., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*, 239-273 doi:10.1017/S030500091400049X

Baltin, M. R. (1989). Heads and projections. In M. R. Baltin & A. S. Kroch (Eds.), *Alternative conceptions of phrase structure* (pp*.* 1-16). Chicago: University of Chicago Press.

Bannard, C., Lieven, E., & Tomasello, M. (2009). Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences, 106*(41),17284‑17289.

Bloom, L. (1970). *Language development: form and function in emerging grammars.* Cambridge, MA: MIT Press.

Brown, R. (1973). A *first language: the early stages*. Cambridge, MA: Harvard University Press.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

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

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge: C.U.P.

Christiansen, M. H. & Chater, N. (in press). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral & Brain Sciences.* Available on

[http://www.ncbi.nlm.nih.gov/pubmed/25869618](https://mail.savion.huji.ac.il/owa/redir.aspx?C=3UOZizaOUEect3tKkSTQhqNleTH9g9MIHiT7FbNpBvOk-52irhTim4kqaiQ6fjF6upgjeaFK-iQ.&URL=http%3a%2f%2fwww.ncbi.nlm.nih.gov%2fpubmed%2f25869618))

Epstein, S. D., Groat, E. M., Kawashima, R., & Kitahara, H. (1998). *A derivational approach to syntactic relations*. Oxford: Oxford University Press.

Erickson, L. C. & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review, 37*, 66-108.

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.

Heny, F. (1979). Review of N. Chomsky (1975), *The logical structure of linguistic theory*, New York: Plenum. *Synthese, 40*, 317-352.

Hofstadter, D. R. (1999). *Gödel, Escher, Bach: An eternal golden braid*. New York: Basic Books. (Originally published in 1979).

Hudson, R. (1990). *English Word Grammar*. Oxford: Basil Blackwell.

Hudson, R. (1995). *Word meaning*. London: Routledge.

Hudson, R. (2014). *An encyclopedia of Word Grammar and English grammar.*  http://www.phon.ucl.ac.uk/home/dick/enc2010/frames/frameset.htm

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. (1982). Basic syntactic processes. In S. A. Kuczaj (Ed.), *Language development 1. Syntax and semantics* (pp. 73-136). Hillsdale, NJ: Erlbaum.

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

Ninio, A. (2006). *Language and the learning curve: A new theory of syntactic development.* Oxford: O.U.P.

Ninio, A. (2014a). Learning a generative syntax from transparent syntactic atoms in the linguistic input. *Journal of Child Language, 41*, 1249-1275. doi:10.1017/S0305000913000470

Ninio, A. (2014b). Variables and values in children’s early word-combinations. *Psychology of Language and Communication, 18(2),* 106*-*125*.* doi: 10.2478/plc-2014-0009

Powers, S. M. (2002). Merge as a basic mechanism of language: Evidence from language acquisition. In E. Witruk & A. D. Friederici (Eds.), *Basic functions of language, reading and reading disability* (pp. 105-117). Dordrecht, Netherlands: Kluwer.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language.* London: Longman.

Radford, A.(1990). *Syntactic theory and the acquisition of English syntax.* Oxford : Blackwell.

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.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge: C.U.P.

van Langendonck, W. (1987). Word Grammar and child grammar. *Belgian Journal of Linguistics,* *2,* 109-132.