First Language, 14 (1994), 119–152. Printed in England

Predicting the order of acquisition of three-word constructions by the complexity of their dependency structure*

ANAT NINIO, *The Hebrew University, Jerusalem*

ABSTRACT

It was hypothesized that the complexity of the dependency structure of different 3- and 4-word constructions predicts the order of their acquisition. Analysis of an English-speaking child's first 102 sentence types of more than 2 words strongly supported this prediction: the earliest sentences rarely involved constructions in which members of a dependency pair were not immediately adjacent to each other. The findings suggest that dependency theory could provide a viable alternative to phrase-structure grammars as a framework for the study of language acquisition.

INTRODUCTION

One of the central tasks of a theory of language development is to account for the order in which children master different types of verbal constructions. In particular, theories of the acquisition of syntax should be able to provide a motivated definition of syntactic complexity or difficulty, to be evoked as the explanation for the order in which children productively control various sentence types.

This paper presents an attempt to account for the order of acquisition of different three-word constructions by a child learning English as her first language. The relative complexity[1] of such constructions is derived

---

* This research was supported by a grant from the Israel Foundations Trustees. Portions of the research reported here were presented at the Biennial Meeting of the Society for Research in Child Development, New Orleans, March, 1993, and at the Sixth International Congress for the Study of Child Language, Trieste, Italy, July, 1993. Address for correspondence: Department of Psychology, The Hebrew University, Jerusalem 91905, Israel.

[1] The exploration of structural complexity as a determinant of the order of acquisition of different constructions does not mean to imply that it is the sole factor influencing their developmental sequencing. In particular, it is very

0142-7237/94/1402-0119 $5.00

from a type of linguistic theory not previously employed as a framework for developmental research, namely, dependency grammar. The change of linguistic framework was motivated by the apparently insurmountable difficulties faced by theories based on phrase-structure theory, to account for the initial stages of multiword speech in children.

Phrase-structure theory

In the last 25 years, the theoretical framework for the study of the development of grammar in children has mainly been the linguistic theory associated with Noam Chomsky, namely, 'phrase-structure' theory. Researchers into child language have used several different grammars of this type as their linguistic framework, among them Chomsky's Standard Theory (1965), his Theory of Government and Binding (1981), or else Bresnan and Kaplan's Lexical-Functional Grammar (Bresnan 1978).

According to this theory, syntactic structure is built up of part-whole relations between the various sized units to which a sentence may be decomposed. Words of a sentence are seen as grouped together to form intermediate constituents or phrases; phrases in turn group together to form higher level phrases, and, ultimately, sentences. In formal terms, the basic ordering relation acknowledged by such grammars is one of containment; the mathematical object representing the syntactic structure of a sentence is a tree in which the lowest nodes are terminal symbols (standing for words), whereas all the rest of the nodes are non-terminal symbols standing for phrases of increasing breadth. Syntactic rules specify the legitimate configuration of lower-order elements to which a node may be decomposed, specifying simultaneously the composition as well as the linear order of the elements. Thus, the ordering of the terminal nodes (the 'leaves') of a phrase-structure tree provides the linear order of the phonetic or orthographic string which is the sentence actually produced.

Phrase-structure grammar and early word combinations

In general, children are not thought to operate with phrase-structure rules right from the onset of language use. With the exception of some

possible that in addition to structural complexity, various aspects of the linguistic input such as the relative frequency of the constructions also have an effect on the order at which they are acquired; see for example Nelson's experimental and correlational studies on the subject (Nelson, Carskaddon & Bonvillian 1973, Nelson, Denninger, Bonvillian, Kaplan & Baker 1984). The examination of such input effects is however beyond the scope of the present paper.

very early work in this tradition (e.g., McNeill 1966), it is generally acknowledged that children's earliest, mostly two-word, combinations cannot be described in terms of a constituent-structure grammar, as there is no evidence that they are generated in terms of constituents of a higher order than words. Such early word combinations are thought to be generated on the basis of pre-grammatical principles which are variously said to be communicative, *ad hoc*, semantic, or purely conceptual, rather than properly speaking syntactic (e.g., Atkinson 1985, Chomsky 1975, Peters 1986, Pinker 1984).[2] True grammar is thought to emerge only at a slightly later stage, when children's constructions – typically of three words or more – start to include what in constituent-structure analysis are noun or verb phrases. At this stage in development, it is claimed, children start to employ for the first time their *innate syntactic knowledge* which is 'triggered' and becomes operative at this point as a consequence of maturation, or of the child's having been exposed to linguistic input in his or her specific language (e.g., Pinker 1984).[3]

Thus, phrase-structure oriented nativist theories of acquisition distinguish between two discontinuous stages in children's production of multiword utterances.[4] The first, pre-grammatical stage is thought to precede the triggering of innate syntax, and the second, grammatical stage to follow it. The implicit grammatical rules underlying the generation of word combinations after the onset of true grammar are not seen as developing from, or constituting elaborations of, the rule system employed prior to this stage. The latter are seen as a kind of

[2] A similar claim has been recently made by Tomasello (1992: 226-228) who has analysed two-word constructions in terms of Cognitive Linguistics rather than a Chomskian phrase-structure grammar. Tomasello disregards the fact that his subject did mark different grammatical relations, e.g., of subject-predicate and of object-predicate, by a different word order, as he believes this constitutes the mere mimicking of adult ordering preferences rather than a true grammatical device to mark the nature of the grammatical relation.

[3] Not all nativist theories posit discontinuity in development. For example, Valian (1986, 1991) claims that development consists of the gradual removal of performance limitations hindering the initial expression of children's competence. Borer & Wexler (1987) also think that development is essentially continuous, maturational asynchrony explaining the later emergence of some types of constructions otherwise expected earlier on theoretical grounds. However, the mainstream of nativism tends to follow Chomsky (1975) and treat the earliest word combinations as essentially discontinuous with later syntax.

[4] Apart from the question whether such a discontinuity is supported by actual child speech data, the result of this move is that children's earliest word combinations, the very data set that because of its simplicity reflects its structure in the most transparent way, is not to be accounted for by the developmental theory.

temporary 'crutch' to make communication possible; or else, a developmental 'blind alley' which children abandon as soon as innate knowledge becomes available to them. True syntax – at least as it is reflected in children's language production – starts only with three-word combinations.

The order of acquisition of different three-word sentences according to phrase-structure based theories of acquisition

The main predictions of phrase-structure oriented, nativist theories of acquisition regarding the order in which different three-word constructions emerge in child language are ones of simultaneity rather than ordering.[5] First, under nativist assumptions concepts representing higher order constituents such as NP or VP are innate; once a child encounters the right kind of environmental output, an equally innate 'triggering' process activates the relevant concept (e.g., Pinker 1984: 37-42). From that moment on, that concept is available for the child in order to serve as a component in syntactic rules he or she may acquire. Second, rewrite rules – the basic form of syntactic rules in a phrase-structure grammar – are context-free substitution rules enumerating the legitimate configuration of lower order constituents to which a higher order symbol can be decomposed. It follows that once a child masters, e.g., the context-free rule NP —> Det + N, this rule should be utilizable immediately in sentence construction for any NP, whatever its role in the sentence.

However, a review of the literature reveals that the prediction of simultaneous emergence for identical expansions of a non-terminal symbol in different grammatical roles in the sentence is very often contradicted by the empirical data. For instance, in many children acquiring English, various expansions of the subject NP to Det + N, e.g., 'This glass broke' or 'My foot hurts', do not appear until later than the formally equivalent expansion of the object noun-phrase as in 'Take this apple' or 'Take my hand' (Bloom 1970, Bloom, Lightbown & Hood 1975, Brown 1973). In some nativist theories a hypothetical performance limitation is evoked to explain these findings, namely, that

[5]. There are several different nativist theories of language acquisition; the following text treats Pinker's model (e.g., 1984) as prototypical or mainstream nativism. Other nativist theories (e.g., Hyams 1987, Kazman 1988, Radford 1990) may not make the same predictions of simultaneous emergence as Pinker's; for example, while Hyams (1987) expects an asymmetry between the use of subjects and objects, Kazman (1988) does not; and Radford (1990) claims that children do not have either subjects or objects in early speech, properly speaking.

it is more effortful to generate 'heavy' or expanded constituents at the beginning of a sentence than at its end, as during the processing of these elaborated constituents the planning of the rest of the sentence or the possibilities of integration into it of earlier units must be kept in working storage (Pinker 1984: 131-133, and see also Valian 1986, 1991: 32). As a claim about a general performance limitation this is probably incorrect; children do not seem to find all 'heavy' initial phrases equally difficult to produce. For example, at the period when Brown's subject, Adam, did not yet produce expanded subject NPs except sporadically, he did produce expanded initial VPs such as 'drive up here', 'put up dere', or 'take off here' (see Pinker 1984: 143). Similarly, Bloom *et al.*'s (1975) subject Gia had, at observation II, a productive verb-particle-object pattern, much before she started to expand subject NPs. A general processing difficulty blocking expansion at the beginning of sentences should apply across the board; as it seems not to, it cannot be evoked to save the hypothesis of the simultaneous emergence of identical symbols from the counter-evidence of the subject-object asymmetry in production.

Another type of discrepancy between the predictions of phrase-structure oriented acquisition theories and developmental evidence concerns the emergence of sentences involving verbal complement-taking verbs such as 'want see picture'. Analysed by the tools of constituent-structure grammar, such sentences are very complex constructions, involving subject-equi deletion or its equivalents. Equi-deletion is a hypothetical process that deletes subjects of complement verbs when they refer to the same individual as the subject or the object of the higher or 'matrix' verb. This process is thought to produce the obligatory subject-less infinitive complements of verbs such as want, try, promise, order and the like, as in, e.g., 'I wanted him to leave', 'I tried to open it', or 'I promised John to return his book by tomorrow.' Another and even more complex transformational process called subject or object raising is posited to produce sentences such as 'John seemed to like Mary' or 'I believed him to be a good friend'. Here, according to theory, the subject of the infinitive complement 'to like' or 'to be' is removed from its original syntactic role and placed into the subject or direct-object slot of the higher verb, in this case, of 'seemed' or of 'believed'. On any theory that relates order of acquisition to complexity, such constructions should emerge very late in child language. In fact, however, constructions involving 'want', 'see', 'can' and other verbal complement-taking verbs appear *early* in child speech, rather than late, including V+V+O combinations such as 'want see picture'. Such findings create an unsolved problem for the constituent-structure oriented acquisition theory (Pinker 1984: 215).

In summary, phrase-structure theory oriented, nativist theories of the acquisition of syntax do not succeed very well in predicting the order of emergence of different three-word combinations in child language. In the following, an attempt will be presented to predict the same phenomenon by the tools of a different grammatical theory, namely, Dependency theory.

Dependency grammars

Phrase-structure theory and dependency theory are two different methods of conceptualizing the linguistic structure of sentences. In grammars constructed within dependency theory (e.g., Hudson 1990, Mel'cuk 1979, 1988), syntax is handled in terms of grammatical relations between pairs of individual words, such as the relation of the subject to the predicate or of a modifier to a common noun. Grammatical relations are seen as subtypes of a general, asymmetrical *dependency relation*: one of the words (the *head*) exhibits a host of local control phenomena towards another word (the *dependent*). First, heads determine the syntactic and semantic features of the head-dependent combination, so that in most cases the combination inherits the features of the head-word. For example, a modified noun such as 'white cat' is still a noun just like 'cat', as far as its semantics and its syntactic combinatory behaviour are concerned. Second, heads control the characteristics and placement of their dependents: for example, the transitive verb 'hit' in 'John hit Fred' requires a pre-verbal subject nominal complement ('John') and a post-verbal direct object ('Fred'). According to theory, the syntactic structure of a sentence as a whole is built up from such pairwise dependency relations between individual pairs of words.

Formally, the dependency relation is an antisymmetrical, antireflexive and antitransitive ordering relation. Two restrictions are placed on the dependency structure of a grammatical sentence: first, every word (but one) must have a head, and second, every word has only a single head.[6] The exception is the *root*, namely, the highest word of the sentence, which does not have a head. Although the number of heads per dependent is restricted to one, the converse is not true so that a head may have a theoretically unlimited number of dependents. For

[6] Some dependency analysts allow multiple heads (e.g., Hudson 1984, 1990), but most do not. Multiple heads create intractable mathematical structures for sentences as well as being unnecessary for the efficacy of dependency analysis; they will not be adopted in the present work.

example, a verb may take multiple adverbial adjuncts, as in 'Mary lives in London near the British Museum on the second floor of a renovated apartment house.'

The overall syntactic structure of a sentence is obtained by the putting together of the different dependency relations existing between its elements. In mathematical terms, the dependency relation imposes a hierarchical structure on the words of a sentence which has the characteristics of a directed tree. A directed tree is a completely connected, two-dimensional, directed acyclic graph with a single root. Each node of the tree represents a word and directional arches between the nodes represent the dependency relation, leading from head to dependent. The tree is headed by the highest word in the sentence, the root, which is the word that does not possess a head of its own.

Apart from the syntactic tree-structure, dependency theory also acknowledges another structural representation of a sentence: that of the sentence as a phonetic or orthographic string. The sentence-string consists of the words of the sentence ordered in sequence, when adjacent pairs of words stand in a relation of being a *neighbour* of each other. The resultant structure is linear or one-dimensional.

According to dependency theory, the transferral from the two-dimensional tree-structure of a sentence to its one-dimensional phonetic chain-structure is regulated by a single rule of the grammar, that of adjacency or projectivity (Mel'cuk 1988, Robinson 1970): a dependent must appear in a sentence immediately adjacent to its head except that the two may be separated by dependent(s) of either words.[7] This rule is applied recursively so that if the inserted dependent has a dependent of its own, the latter may in turn be inserted between its own head and *the head's* head. The result is that the *deeper* the dependency relation between a pair of words on the tree (the further away from the root), the *closer* are the words in the linear sentence. In other words, the degree of proximity of two grammatically related words in the sentence-chain encodes the relative subordinate-ness of the relevant dependency relation, in comparison to other relations either word participates in. The technique is similar to that employed in drawing in perspective: the further away a figure from the viewer in the three-dimensional space represented by the picture, the smaller it is drawn in the picture-plane.

---

[7] Hudson (1990) proposes a modified adjacency principle by which a dependent may also be separated from its head by a word which is the mutual head of both. However, this modification is necessary or relevant only in a system that allows multiple heads.

In summary, dependency grammar characterizes syntactic structure in the main in terms of dependency relations among the words of a sentence, without resorting to units of analysis smaller or larger than the word. In particular, combinations of words are not represented formally in the grammar in the form of independent symbols on which grammatical rules are predicated. The head of each pairwise combination represents the combination as a whole in all further dependency relations; phrases and clauses are thought to be redundant as formal linguistic entities as all that needs to be posited about such non-terminal constituents can be posited about their highest word without loss of information. Informally, it is possible of course to speak of phrases, clauses or sentences; it is only as formal components of syntactic rules that these are deemed redundant. The only place in the grammar when constituents larger than words are used is in co-ordination; but even then the conjuncting units are not phrases in their phrase-structure definitions. For details, see Hudson (1990: 405). In the same vein, it is possible to speak informally of words as constituents of combinations, whether of phrases, clauses or sentences; however, the syntactic rules regulating the lawful creation of structure do not make formal use of a containment relation between larger units and their constituents. The place of 'rewrite rules' familiar from phrase-structure approaches, specifying the 'daughters' of a phrasal symbol, is taken by specifications of the direct government or dependency relation between 'sisters'. Dependency grammar works with horizontal rather than vertical grammatical relations.[8]

Early word combinations in dependency-oriented acquisition theory

In contrast to some phrase-structure oriented acquisition theories that posit a sharp discontinuity in children's use of language before and after the appearance of higher-order constituents in children's speech, a dependency-theory oriented acquisition theory can, potentially, view all of syntactic development as a continuous process, beginning with the earliest two-word combinations. Whereas it is extremely difficult to apply phrase-structure analyses to children's two-word combinations, the converse appears to be true for a dependency-type analysis. As has been pointed out repeatedly (e.g., by Bloom 1970, Bowerman 1973,

[8] In his latest publications, Chomsky (e.g., 1992) appears to be moving towards a theory where 'horizontal' grammatical relations are accorded an increasingly prominent role in the grammar, and 'vertical' relations a relatively attenuated role. There is a distant possibility that phrase-structure and dependency theories will eventually fuse into a unified syntactic theory.

1976, Braine 1976, Brown 1973, Ingram 1989, Ninio 1988, Tomasello 1992), early multiword utterances are built around a random collection of 'predicate words' which children combine with words representing their arguments in the relevant propositional function. The structure of such early utterances is usually described in semantic terms, for instance as the combination of function words or relational words and their arguments, but it is easy to see that, in terms of a dependency grammar, these word combinations are really syntactic ones. Given a dependency approach, they can be said to demonstrate children's having acquired the dependency and linear ordering relations between pairs of words, one of which is a headword, the other, one of its semantic/syntactic dependents.

At this early stage of generating multiword utterances, involving as it does only a single dependency pair, learning is restricted to the mastery of the combinatorial properties of 'predicate words' and of the ways in which a given predicate-argument relation is to be realized in a correctly ordered sequence. Theoretical considerations as well as empirical findings suggest that this learning is most likely to be lexically specific and item-based, involving each predicate word separately (e.g., Bowerman 1976, Braine 1976, Ninio 1988, Tomasello 1992). These individual dependency relations are the fundamental building blocks of a dependency structure. At this stage, apparently nothing else is learned: children do not need either to co-ordinate several different dependency relations or to master the intricacies of the Adjacency Rule. The acquisition of syntactic knowledge starts with mastering the single syntagmatic couple.

It is, of course, logically possible that phrase-structure oriented theories of language acquisition are correct, and that even though children may start the process of acquisition by acquiring the rudiments of a dependency grammar, as soon as innate syntactic knowledge becomes available for them, they abandon dependency grammar in favour of a phrase-structure type one. However, the discontinuity claim is not supported by developmental evidence: the longer, supposedly 'grammatical' constructions can be and have been shown to emerge in smooth continuity and in small incremental changes from the supposedly 'pre-grammatical' shorter constructions (e.g., Brown 1973, Elbers 1990, Ewing 1982, Hill 1984, Ninio 1988, Schlesinger 1982, Tomasello 1992). Rather than exhibiting a sharp discontinuity, children's grammatical development shows the signs of a continuous process, albeit not in terms of a phrase-structure grammar but in terms of a dependency-structure one. Children at the onset of acquisition appear to acquire a dependency syntax.

If this conclusion is correct, it should be possible to predict the order of acquisition of different three-word constructions on the basis of their relative difficulty in terms of their dependency structure, regardless of their characterization in terms of their constituent structure. Such a prediction forms the hypothesis of this study.

The complexity of three-word combinations in terms of their dependency structure

According to the analysis presented above, children start the acquisition of syntactic knowledge by learning to express, in a two-word utterance, the syntagmatic relation existing between a pair of words. The developmental history of this achievement suggests that various components of this novel verbal schema may be learned separately. In particular, it has been observed that potential members of syntagmatic pairs start to appear in children's single-word speech several weeks or months before the generation of true two-word utterances, in the form of intonationally independent successive single-word utterances (Bloom 1973, Horgan 1976, Ingram 1979, Scollon 1976). The production of integrated two-word utterances takes a further period of learning in which the pause between the words is gradually reduced until finally they are brought under a single shared intonation contour (Branigan 1979, Fonagy 1972, Scollon 1976). The significance of this development in dependency-grammar terms is that, at this stage, children learn how to produce two syntagmatically related words as *neighbours* in a vocal string. The neighbourhood or adjacency relation is linguistically as well as psycholinguistically of central importance. Linguistically, adjacency (or juxtaposition) is the central and sometimes the sole marker of two words being in a dependency relation (e.g., Goldenberg 1989, Noonan 1985). Moreover, the syntactic rule regulating the linear ordering of words in a sentence, namely, the Projection Principle (Robinson 1970), is defined solely in terms of the precedence of lower relations over higher ones, to be neighbours in the sentence chain.

In terms of the psychological processes through the means of which sentence generation and interpretation are actually carried out, the neighbourhood in the vocal string of two words which are syntagmatically related means that these may be accessed in immediate succession. In psycholinguistic terms, a dependency relation is a computational command to the effect that during sentence generation and comprehension, the information carried by the two separate words comprising the dependency couple is to be combined or synthesized; if these words are immediately consecutive, this procedure can be carried out without recourse to storage in, and retrieval from, short term memory. During sentence generation, if the two members of a dependency couple are to be separated by some intervening material in the sentence, the processing of one dependency relation is interrupted by the processing of another, thus creating *an open dependency* for the duration. Until the second member of the couple is generated, the speaker has to keep in short-term memory the fact that such a closure is pending. For example, the generation of 'Give him apples' requires that the speaker store the demand for the second dependent of 'give' until after the first dependent ('him') is produced.

Open dependencies also demand extra processing during the interpretation of sentences by listeners. They have to store in memory all words whose requirement of a head have not yet been satisfied until the head word is encountered (e.g., in the sentence 'This boy came', the word 'this' which is considered the head of the following noun, is separated by it from its own head, the tensed verb 'came'). In addition, they have to backtrack their way in the sentence when they encounter a word whose head is a word that has already been processed (e.g., 'apples' in 'Give me apples', or 'quickly' in 'Come here quickly'). There is evidence from adults that the storage and retrieval processes involved take up more than ordinary short-term memory: there seems to be a very strict absolute limit on the number of open dependencies operative at the same time in a sentence, so that if their number passes three, the sentence is incomprehensible. Sentences with three simultaneously open dependencies are already extremely difficult to process, for example, 'Who did the dog which the farmer who the cat licked owned chase?' (Hudson 1992, 1993).

The linguistic and psycholinguistic approaches to the role of adjacency lend themselves to mathematical formalization. The sentence-chain is a one-dimensional object, a linear section or part of a line, in which the ordering relation is one of neighbourhood. In graph-theoretical terms (e.g., Tutte 1984), the question is whether a given section (or verbal string) can or cannot accommodate a set of different ordering relations – namely, dependency relations – among its elements. The answer is that the linear section can accommodate the set of dependency relations if and only if each dependency relation coincides with a neighbourhood relation between the elements of the string, namely, if all dependency couples are also neighbours in the string.

Translated to processing terms, this means that a non-adjacent dependency requires a detour from linear processing into and out of memory storage, and that the combinatorial operation cannot be performed without recourse to such storage. In linguistic terms, the implication is that unless dependencies coincide with neighbourhood, there is a need to evoke the priority principles of the Projection Rule in order to place correctly (or identify) dependency couples with conflicting rights to adjacency. If however the linear structure can accommodate the dependency relations, there is no need to evoke the Projection Rule to interpret or generate the sentence.[9]

It follows that there is a very clear mathematical criterion to distinguishing between more and less complex three-term combinations. Those that accommodate the dependency structure of the sentence in the linear structure of the verbal string are less complex than those that do not; it can be predicted that the former will be acquired earlier than the latter.

Table 1 presents all possible types of well-formed three-term combinations, according to their dependency and linear structure, with some illustrative examples for each pattern. For expository purposes only, a letter-based notation is used to characterize the abstract structure of the different patterns. The words of the pattern are represented by the letters H (for head of a dependency pair) and D (for a dependent). If a head has two dependents, these are marked D_1 and D_2. If a dependent has a dependent of its own, the subordinate dependency couple is put in parenthesis and its members are designated simply as H and D, without otherwise noting that the head of the subordinate combination is also the dependent of its own head, the H outside the parenthesis. Thus, (H D) H is to be read as a construction in which the *third* word of a sentence is the head of the first word, which in turn is the head of the *second*.

Even though the H-D notation is sufficient to represent the dependency structure of a three-word combination, a redundant notation is also used in the form of directed arrows drawn below these symbols. For each dependency couple, the direction of dependency is marked by an arrow from the head to the dependent. For example, in the first pattern of Table 1, the structure, in the abstract, is noted H (H D), and the arrow configuration (duplicating this information) is –>–>. Such a notation is to be read as indicating that the first word of the sentence is the head of the second, and the second is the head of the third.

[9] An alternative formulation of the mathematical features of non-linear structures is that the *path* defined on the structure of a non-adjacent pattern is *re-entrant* or circular rather than simple. This definition emphasizes the fact that after processing one dependency pair, the processor needs to return to an already processed element in order to form another dependency couple with it.

TABLE 1. *Types of 3-word constructions by their dependency structure**

A. Linear constructions – all dependency pairs adjacent		
#1	H (H D) —> —>	Take this ball
#2	D_1 H D_2 <— —>	John made this
#3	(D H) H <— <—	Little boys danced
B. Non-linear constructions – one dependency pair non-adjacent		
#4	H D_1 D_2 —> ———>	Put it there
#5	D_1 D_2 H <——— <—	This I dropped
#6	H (D H) ———> <—	Bring big apples
#7	(H D) H —> <———	The boy disappeared

* H: Head
D, D_1, D_2, (H): Dependent(s)
H —> D: Direction of dependency

The examples in Table 1 represent only one of a large number of different 3-term combinations possessing the relevant dependency structure. Thus, the first pattern, i.e., —> —> , while representing the structure of a sentence such as 'take this ball', also characterizes utterances such as 'in the house' or 'want see picture'.[10] Similarly, the second pattern, i.e., <— —> , represented by the subject-verb-object combination of 'John made this' in the Table, is also the structure of sentences such as 'John run fast', where the last word is an adverbial adjunct of the verb 'run' rather than its direct-object complement.

It is immediately evident that the first three patterns of Table 1 are those in which the dependency relations coincide with the neighbourhood relations between the words. The other four patterns are those in which one of the dependency couples is not adjacent in the sentence and therefore a single linear graph cannot accommodate the

[10] In the analysis adopted in this project, determiners are considered heads of the common noun complementing them. See Method for further discussion.

dependency structure of the sentence. It is predicted that three-word combinations whose dependency structure can be represented by a linear graph will be acquired earlier than those whose structure is to be represented by two-dimensional graph.

It is noteworthy that the distinction we have made between the one-dimensional and the two-dimensional structures is orthogonal to a classification made solely on the basis of the abstract syntactic structure of the same three-word combinations. An abstract representation in the form of a syntax tree of the dependency structure of three-word combinations distinguished between two patterns only: one a branching tree in which there is one head with two dependents, and the other a non-branching structure in which the (single) dependent of the first relation has a (single) dependent of its own. Patterns #2, 4 and 5 in Table 1 are the branching kind, whereas patterns #1, 3, 6, and 7 are the non-branching kind; namely, both abstract syntactic types are represented in the linear as well as the non-linear categories. In other words, in the present approach the syntactic structure and thus the complexity of sentences is not characterized only by their abstract organization but by the *interaction* of dependency and adjacency considerations.

This difference in approach explains why a dependency-oriented acquisition theory makes such very different predictions about the relative order of mastery of different three-term constructions, relative to the predictions of a phrase-structure oriented theory like Pinker's (1984). Patterns which under phrase-structural assumptions are expected to emerge at the same time as they make use of identical context-free rewrite rules, create, in dependency-structural terms, 3-term combinations of differing structural complexity, and therefore are expected not to emerge simultaneously. In particular, under dependency analysis, identical expansions of higher-order constituents do not add the same amount of structural difficulty to a sentence independent of their grammatical role in the sentence. For instance, the expansion of an object NP to Det + N creates a simpler 3-term dependency-structure (#1) than the identical expansion of a subject one (#7).

Other possible determinants of complexity

Apart from issues of adjacency, a dependency analysis suggests several other factors that may influence the relative difficulty of a three-word construction, among them the type of grammatical relations involved, the number of dependents per head, the direction of the dependency relations relative to the dominant direction of most dependency relations in the language, and more. However, in contrast to the factor of open dependencies for which a well-reasoned case can be made, the effect of these other factors on the developmental ordering of 3-term constructions is a matter of speculation. Intuitively, it seems obvious that some grammatical relations are more complex than others, in terms of the information manipulation involved. For example, modification (as of the noun by the prenominal adjective in 'white cat') is probably more complex than complementation (e.g., of the transitive verb by its subject and direct object in 'John saw Mary'). Some words combine with others in especially complex ways, for instance, the syntactic behaviour of relative pronouns or *wh*-elements is more complex than of e.g., common nouns.[11] Even though this factor is not systematically investigated in the present study, it can be predicted that the first 3-words sentences generated by a child will comprise of relatively simple grammatical relations, the chief among them complementation, rather than of more complex ones.

Similarly, it is possible that the co-ordination of two dependents for a single head is more complex than the chaining of several single-dependent relations, as the former involves some extra lexical learning of the complex argument structure of multiple-argument predicates. However, it is difficult to weight this factor against the increased complexity of a three-level hierarchical structure created by chained dependencies.

Lastly, it is well known that most languages of the world, and among them English, have a dominant word-order for the placements of heads and dependents. For instance, Welsh is a language where heads usually precede their dependents, whereas Japanese is a strongly head-last language (Hawkins 1983, 1990, Siewierska 1988). English is on the whole a head-first language, with the exception of the placements of subjects, prenominal modifiers, and some other minor patterns (Hudson 1990).[12] It seem intuitively reasonable that children would find it easier to learn to generate constructions in which the dominant order of the head-dependent relation is conserved. However, it is also possible that children do not generalize head-dependent ordering rules across different types of grammatical relations and therefore this factor does

[11] A formal introduction of the information-combining operations involved in different grammatical relations is outside the scope of this paper. Intuitively, modification is more complex than complementation because the former involves the logical Lambda Conversion whereas complementation involves straightforward constant-for-variable substitution (Dowty, Wall & Peters 1981, Montague 1974).

[12] In a large sample of English utterances, consisting of about 100,000 head-dependent pairs, about 75% were head-first (Hudson 1990:212).

not have an effect on the relative difficulty of patterns with head-initial, mixed or head-last dependencies. Without empirical evidence on acquisition patterns, the role of these factors in determining developmental order cannot be established with any confidence.

To summarize, the main hypothesis of this study is that it is possible to predict the order of acquisition of different three-word constructions on the basis of their relative difficulty in terms of their dependency structure, regardless of their characterization in terms of their constituent structure. It is expected that the major determiner of complexity will be the linearity of the constructions; namely, that the earliest combinations will consist of patterns in which all dependency couples are immediately adjacent. Since linearity as defined above cannot be derived from a phrase-structure based analysis of three-word combinations, the alternative hypothesis under a phrase-structure framework is that linearity will have no effect on the order of acquisition.

METHOD

The language corpus

The data base consisted of the first 102 sentence types of Tomasello's (1992) subject Travis which were of more than two words; they were produced when Travis was between 1;6.8 and 1;7.18.[13] For comparison, the next 272 sentence types of more than two words were also analysed; these were produced between 1;7.19 and 1;8.8. This age is the limit of the parents' recording every sentence of Travis; after this age utterances were recorded only if their structure was more complex than of those produced previously. The source for the analysis were Table 8.2 and the Appendix of Tomasello's (1992) book *First Verbs*.

In defining words, Tomasello's decisions about compounding were followed, e.g., 'stop-it', 'get-it', 'up-there' and 'right-there' were considered single words. In addition, two-part proper names such as 'Peter Pan', 'Sesame Street' or 'Captain Book' and some compound object names such as 'paper-towel' and 'rubber-band' were also considered single-word expressions, whether or not they were compounded in the original. If the child used a compound name

[13] Originally, the arbitrary number of the first 100 types was chosen to represent the initial period of acquisition but the observation in which the 100 mark was reached contained 2 extra sentences which were thus also included in the data base.

contrastively, as was the case for instance with 'banana popsicle' and 'orange popsicle', these were considered two-word expressions, the first a modifier of the second.

As this corpus was collected in order to investigate the child's acquisition of non-nominal forms, object labelling was not systematically recorded and such utterances were excluded by Tomasello from the reported corpus. The exceptions were complex naming sentences, for example ones including a copula or an adjective, which were included in the reported data base regardless of their use in labelling. Moreover, articles such as 'the' and 'a' (i.e., schwa forms) were also not recorded systematically. This means that the corpus does not include any copula-less naming expressions of three words which the child may have generated before she started to use the copula at around 1;11. Extrapolating back from the copular forms reported for this later period (p. 155), these may have been pivotal combinations with a sentence-initial 'it', 'this', 'that', 'there' and 'here', such as 'That a flower'. Consequently, the relevant patterns are underrepresented in the analysed data base, affecting mostly the proportion of Pattern #1.

Dependency analysis

Dependency relations were identified according to theoretical criteria for identifying heads and dependents (e.g., Gazdar, Klein, Pullum & Sag 1985, Hudson 1987, 1990, Mel'cuk 1988, Pollard & Sag 1987, Zwicky 1985). In the main, the system followed Hudson (1990), except for not allowing either multiple heads or mutual dependencies. Words were considered heads of their syntagmatic or valency arguments when these complemented them; thus, verbs were considered heads of their subjects as well as direct, indirect and oblique objects. In line with Tomasello's analysis of Travis's early use of particles as verb-like forms (pp. 82–93), a particle appearing without its verb was seen as replacing it as the head of the verb's arguments, e.g., the particle 'on' in the sentence 'On the light', standing for 'turn on the light'. Following current practice, both in the dependency tradition (Hudson 1990) and in GB (i.e., the DP-Hypothesis; see Abney 1987, Speas 1990), determiners were considered heads of the common noun complementing them. In sentences with a missing copula (e.g., 'Raisins in there' or 'Bottle coming too'), the bare copular complement was considered the head of the sentence subject.

Lastly, in word pairs involving modification or the adjunct relation such as verbs and adverbs or common nouns and adjectives, the adjuncts were considered the dependents of the words they modified.

RESULTS

The characteristics of Travis's first word combinations of more than two words

Appendix A presents the first 102 word-combinations of three words or more produced by Travis, categorized according to their dependency structure. Of these, 90 were of three words and 12 of four words. As the great majority of the sentence types were of three words, the results of this initial stage will be presented in terms of 3-element constructions, analogous to Table 1, while the few 4-word constructions she produced will be assimilated to the 3-word pattern closest to them. Table 2 presents the distribution of the first 102 sentence types according to their dependency structure.

With the exception of just 10 sentences, the overwhelming majority (90.2%) of Travis's first 3- and 4-word sentence types were of the linear type, where all dependency couples are immediately adjacent to each other. Among these, about two-thirds were the heads-first chained type H(HD), and about a third, the head-in-the-middle, branching type D_1HD_2. The other 5 possible 3-term patterns occurred with very low frequency in the corpus.

Moreover, there is reason to believe that the actual percentage of zero-open constructions in Travis's language at this period was even higher than estimated on the basis of the reported corpora. As mentioned in the Method section, the corpus does not include the child's labelling utterances if these did not include a copula or at least a modifier. An informed guess based on the copular forms reported in Tomasello (1992: 155) makes it very probable that most of the copula-less naming expressions of three words were pivotal combinations with a sentence-initial 'it', 'this', and 'that', namely, expressions like 'It/this/that a flower' which fall into Pattern #2 (i.e., D_1HD_2), or else combinations with 'here' and 'there' in a sentence-initial position, such as 'Here/there the lion', falling into Pattern #1, namely, H(HD). In either case, the missing naming utterances are zero-open combinations, and had they been included in the corpus, the proportion of zero-open sentence types at this initial period would probably been much closer to 100%.

Under the alternative hypothesis that children acquire a phrase-structure type grammar, the linearity of a word combination is not expected to affect its chance of being acquired as part of the earliest set of multiword sentence types. The results demonstrate conclusively that this hypothesis is untenable; linearity, a dependency-defined measure,

TABLE 2. *Distribution of Travis's first 102 sentences of three words or more by their dependency structure*

Dependency pattern Number*	Structure	Frequency	%
Total linear patterns		**92**	**90.2**
Total Type 1		*65*	*63.7*
1	H (H D) —> —>	64	62.7
1A	H (H (H D)) —>—>—>	1	1.0
Total Type 2		*24*	*23.5*
2	D_1 H D_2 <— —>	15	14.7
2A	D1 H (H_2 D) <— —> —>	9	8.8
3	(D H) H <— <—	3	3.0
Total non-linear patterns		**10**	**9.8**
Total Type 4		*7*	*6.9*
4	H D_1 D_2 —> ———>	5	4.9
4A	H (H_1 D) D_2 —> —> ————>	2	2.0
5	D_1 D_2 H <——— <—	1	1.0
6	H (D H) ———> <—	1	1.0
7	(H D) H —> <———	1	1.0

* Four-word constructions are numbered as the closest three-word pattern, with the addition of an 'A'.

has a decisive effect on the types of constructions first acquired by this child.

Further developments

Of the next 272 sentence types produced by Travis, 190 were of 3 words, and 82 of 4 words or more. Of the total 272 sentence types, 202 or 74.3% were the linear type without an open dependency, the rest, 70 or 25.7%, were the non-linear type. A Chi-square test of the significance of this difference produced a significant result (Chi-square of 1 d.f. was 11.50, $p < 0.001$). With the reservation that patterns #1 and #2 are still underrepresented at this period, the comparison with the first 102 sentence types produced by Travis reveals a considerable increase in the more complex constructions at this second period of the development of three-plus word combinations. Namely, the almost exclusive production of linear sentence types is a feature of this child's initial period of combining three or more words into a sentence, and not a constant of her language use at all ages.

To trace the developmental trend of the acquisition of non-linear, one-open patterns in finer detail, Figure 1 presents the cumulative number of linear and non-linear sentence types per observational session, as well as the total number of sentence types by that observation.

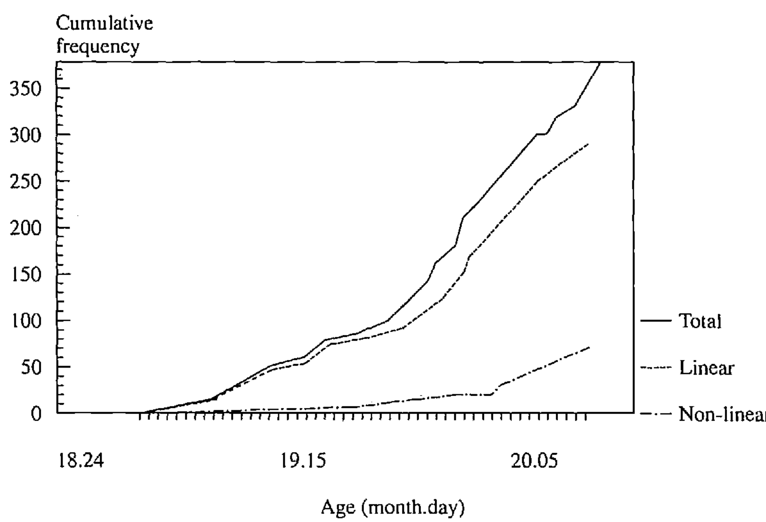

Fig. 1. Cumulative frequency of linear, non-linear and total sentence types, by child's age at observation.

It is evident that for the first month of producing multiword utterances of more than two words, most of Travis's sentence types were of the linear kind. The acquisition of non-linear multiword constructions in any significant numbers started only when Travis already produced about a hundred different linear structures. Till the end of the period covered by the corpus, the rate of acquisition of non-linear patterns remained relatively slow, never approaching the accelerated growth rate of the linear constructions.

In order to investigate the development of non-linear patterns in greater detail, two types of open dependencies were defined. The first was *co-dependent separation*; in these patterns (#4 and 5 and analogous patterns among the longer utterances) a dependent is separated from its head by a word which is also the dependent of the same head. The second was *subordinate separation*; here (patterns #6 and 7 and their analogues among the longer utterances) the dependent is separated from its head by a word which is its own subordinate dependent. To follow up the intuition that co-dependent separation may be a different and less complex operation than separation by a subordinate, the cumulative number of co-dependent and subordinate-separated sentence types was computed. Figure 2 presents the results.

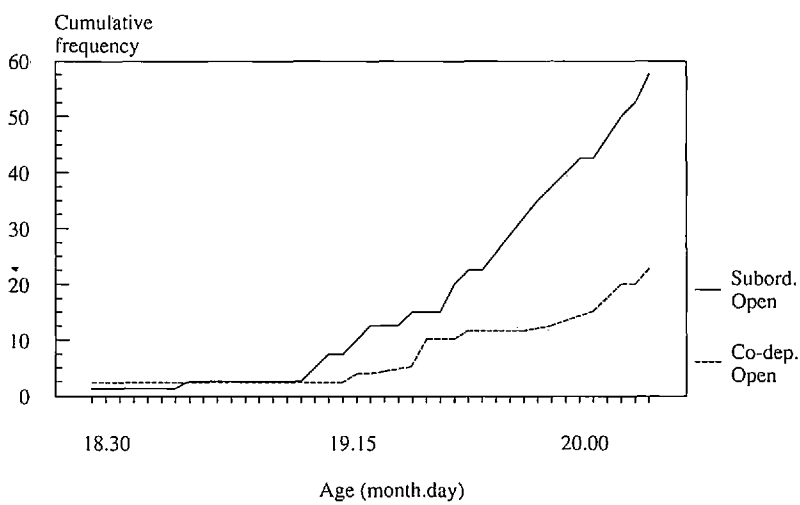

Fig. 2. Cumulative frequency of co-dependent-interrupted and subordinate-interrupted non-linear sentence types, by child's age at observation.

The results show that co-dependent separated-sentence types are mastered earlier and develop at a faster rate than subordinate-separated ones. Namely, the next type of co-ordination appearing after the linear one is co-dependent separation, a type of pattern accounting for less than 8% of the first 102 sentence types.

DISCUSSION

The results of this study revealed that Travis's initial three-term combinations were with few exceptions of the all-adjacent, linear type, and only after a further period of time did the child start to generate non-linear multiword combinations in any numbers.

In terms of the grammatical configuration generated by the child, her preferred initial three-word combinations were of two kinds. One involved a predicate with two valency-complements, the other a main predicate with a single complement that took a complement of its own. The results show that Travis, at the onset of acquisition, learned to generate these relations in a linear configuration, and postponed the generation of sentences with an identical abstract dependency structure but which are realized in a 'non-linear' sentence-chain.

It is impossible to account for this result in terms of a developmental theory that takes a phrase-structure type grammar as the representation of the linguistic knowledge children acquire. Linearity as defined in this study is derived from the dependency structure of sentences; it registers the adjacency of members of dependency couples in the sentence-string. As a phrase-structure analysis of the same sentences does not identify dependency couples, in such a system, linearity (in the sense employed here) is an undefined concept. In the absence of this discriminating feature from the theory, there is no principled way for a phrase-structural grammar to distinguish between sentence types Travis acquired earlier and those she added later to her repertoire; as a prominent example, there is no possible constituent-structural explanation for the fact that some types of four-word sentences appear to be actually easier to master than some fairly simple three-word ones.

In general, constituent-structural, nativist theories lack a theoretical apparatus for recognizing one type of three-word construction as a priori structurally more complex than another, excepting maybe patterns posited to involve deletion or movement such as equi-deletion or raising. The *ad hoc* hypothesis that a general performance limitation accounts for the avoidance of expanded initial elements in a sentence, mentioned in the Introduction (Pinker 1984, Valian 1991), is virtually the only developmental ordering principle arising from this acquisition theory with respect to the relevant developmental period. Pitting this hypothesis against the linearity hypothesis derived from dependency theory as the explanation for the pattern of results obtained in this study, it is obvious that the performance limitation claim cannot explain why Travis did not produce non-linear sentence types (patterns #4, 5, 6) which do not involve a 'heavy' initial load, for example, sentences of the type 'Give me apples' (see also Bloom *et al.* 1975, for more evidence on the late emergence of dative objects). On the other hand, the initial-effort factor does account not only for the asymmetry between determiner-expanded object NPs (Pattern #1) and determiner-expanded subject NPs (Pattern #7) but also for the absence of Pattern #3, not otherwise explained by the linearity factor. This pattern represents the structure, among other sentence-types, of subject NPs expanded by a prenominal adjective. As this pattern represents an initial 'heavy' elaboration, the initial-effort factor predicts that Pattern #3 should be late in acquisition, or more precisely, later than the comparable adjective-noun elaborated object NPs. It is true that Travis did not at the relevant period generate any adjective-elaborated object phrases either (Pattern #6), thus weakening to some extent the impact of this successful prediction, but at the least the general performance factor offers some kind of an explanation why Pattern #3 is missing, even though by dependency analysis it is a linear pattern and should, a priori, be of comparable frequency as Patterns #1 and #2. As we shall see below, there are other possible explanations for the absence of Pattern #3 sentences and ones that take the complete configuration of results into account.

As to the posited explanation in terms of a general processing difficulty, as it has been pointed out previously, it is doubtful if such a non-specific performance limitation accounts for the lack of expanded subject NPs at the onset of three-word speech; other constructions involving an expanded sentence-initial constituent (such as verb-particle-object combinations) are productive in the same children's speech much before they start to expand subject NPs in other than isolated cases (see Bloom *et al.* 1975:50–58; Pinker 1984:143). That does not mean, of course, that a more precise version of this claim, phrased in dependency-grammatical terms, may not be entertained; the proposed factor of the degree of linearity of a particular construction or the type of interruptions in its generation or parsing is precisely that. If so, the linearity feature should also be able to distinguish, as the general performance factor cannot, between early and late expanded sentence-initial constituents; the results show that it indeed does. On a dependency analysis, elaborated initial VPs involve the separation of a dependent from its head by a co-dependent of the same head, whereas expanded subjects – at least those involving determiners – involve the separation of a dependent from its head by its own subordinate. In the present study, co-dependent separation was found in general to be mastered earlier than subordinate separation (see Figure 2), a result expected on a dependency-analytical model of acquisition. It appears that factors sensitive to the dependency structure of sentences – whether we wish to call them factors of competence or of performance – are better predictors of the overall developmental picture than structure-insensitive factors such as a difficulty with increased processing load at the beginning of sentences.

In summary, although the performance-limitation hypothesis has a partial success in accounting for the data, overall, the phrase-structural explanation for the obtained pattern of results is much weaker than the dependency-based one.

As mentioned above, the imbalance among the three different linear patterns in the first 102 sentence-types that Travis initially acquired requires closer examination. As suggested in the Introduction, the disproportion between the head-first chained Pattern #1 and the branching Pattern #2 may be due to the greater difficulty of dealing with multiple dependents than with hierarchically organized chained single dependencies. On the other hand, it is possible that the findings regarding the relative frequency of these two patterns are invalid in the first place, as both of these patterns are in all probability underreported in the data base to an unknown extent. In other words, it is perfectly possible that Travis generated at the relevant period hundreds of naming expressions on the pattern of 'that a lion' or 'that my chair', and had these been included in the corpus, the imbalance between Pattern #1 and #2 would have disappeared or even reversed. It is impossible to determine on the basis of this corpus whether the pattern observed with other than naming expressions represents a true phenomenon of a general preference for non-branching structures in the earliest 3-term combinations; the issue will be solved only when analysis currently under way of the speech of several other children acquiring English is completed.

The situation is different with respect to the difference between the frequency of Pattern #1 and its mirror image, Pattern #3. The latter is apparently not underrepresented in the corpus; such sentences invariably involve a prenominal adjective modifying a common noun, and these were included in the recorded corpus of Travis's language whether or not they were used by her for object labelling. Apart from the putative performance limitation discussed above, there are two different explanations why such patterns were so infrequent at the initial period of 3-word utterances even though they are of the uninterrupted linear type. First, they involve the grammatical relation of modification of a common noun by an adjective; there are good reasons to suppose that this relation, as the adjunct relation in general, is more difficult to acquire than the complementation involved in most sentences in Pattern #1 (see Note 11). Indeed, in the whole period under discussion, there are extremely few multiword combinations with prenominal adjectives, whatever the overall structure of the sentence. However, an explanation of the extreme rarity of Pattern #3 at the onset of three-word speech on the grounds that Travis did not yet know at this stage how to generate modification by an adjective, runs into the difficulty that, by that time, Travis had produced many *two-word* combinations involving prenominal adjectives. Interestingly, a close inspection of the relevant utterances in the corpus revealed that there is no conclusive evidence for these indeed expressing a relation of modification rather than that of predication by the adjective. For example, the sentence ‘Good beer’, said on her stealing a sip from a bottle at 19.04, seems much more likely to be a claim that the beer was good or tasty, than a claim that the beer she sipped was of a good rather than a bad kind. In fact, none of Travis's two-word combinations with an adjective, whether pre- or post-nominal, functioned contrastively, to refer to a specified alternative out of two or more possible referents for the common noun; it is even questionable if the few three-word combinations categorized here as Pattern #3 were used so. If this interpretation is correct, this raises the possibility that Travis did not master adjectival modification until well into the 3-word stage, past her first 100 combinations of more than 2 words. This is an important avenue to explore with respect to beginning speakers in general.[14]

The second possible explanation for the rarity of Pattern #3 in the initial corpus is its involving two dependency relations in which the heads follow rather than precede their dependents. The only other three-term pattern in which this is so is #5, namely D_1D_2H, a pattern occurring only once among the first 100 combinations, and only once later on. The two explanations are not mutually exclusive: in a language which is mostly head-initial, constructions involving two successive dependency relations in the reverse order are unusual and thus will tend to be more complex than the average. Whatever the

[14] Braine (1976:76) made a similar suggestion regarding adjective-noun combinations in the speech of the children he investigated.

precise explanation, the observed rarity of Pattern #3 demonstrates that the other dependency-structural factors besides the linearity of the construction play a role in determining the order of acquisition of different three-word patterns.

Linearity, however, is apparently a very powerful determiner of the syntactic complexity of sentences. As we have seen, it explains – in dependency-grammatical terms – the long-term puzzle in early child language, namely, why is it that English-speaking children generate determiner-common noun sequences in object position earlier than in subject position: the latter represent a non-linear pattern and therefore a higher level of difficulty, in dependency-grammar terms, than the former. Moreover, the same explanation holds for the relative difficulty of interrupted patterns persisting well into mid-childhood: for instance, the asymmetry between subject and object expansions of various kinds (e.g., by a relative clause) is evident until age 7 or more (Menyuk 1969, 1977). Children at this later age also have difficulties mastering subject-control verbs in which the embedded verb is separated from its (semantic) subject by the main direct object (Chomsky 1969). These findings suggest that for any novel syntactic pattern, the linear version will be mastered before the non-linear one. Interruption of processing appears to be a lasting source of difficulty in the generation as well as the comprehension of sentences.

There are several convergent explanations why the mathematically simpler structure is also earlier in acquisition. First, a child can learn to produce a linear three-word combination without having first to master the rules regulating the rights of other words to separate a dependency couple. In order to produce a non-linear three-term combination, the child must be able to balance the conflicting rights of two words to be adjacent to a given word, whether of two dependents of the same head or of the dependent and the head of the same word. Even if the child is willing to produce such constructions without actually figuring out the correct priority rights, the sheer fact that there is a conflict around placement of words in the sentence may slow down the acquisition of such patterns.

Second, linear three-word constructions enable the speaker to complete the processing of both elements of a dependency couple without any interruption. As the precognitive processes involved in sentence generation are not as yet known, it is not clear what is it in the processing of open dependencies that cannot be mastered by a beginning speaker. It is possible that the information-processing procedures to be used for generating non-adjacent dependencies are structurally different from those for expressing the two components of a dependency couple in succession; a beginner may not as yet possess the correct schemas for dealing with interrupted relationships. The non-linear patterns produced by this child in the first month of generating three-word sentences were so few in absolute numbers (5 sentence-types for the most frequent non-linear pattern; 2 or less for four others) that it is very possible that these do not represent a truly rule-governed behaviour at all. In any case, the mathematically simple structural pattern represents a less complex procedure for generating three-word sentences, and it appears that the simplicity of procedures involved in generating sentences is a decisive predictor of the order of their acquisition. If this conclusion is correct, this is important support for a conceptualization of syntactic rules in procedural rather than declarative terms.

This notion receives further support from the finding that past this initial period of acquisition, the next type of combinatorial procedure mastered by Travis is that of the separation of a dependent from its head by a co-dependent, whereas separation by a subordinate of the dependent requires a yet further stage in acquisition. The developmental differentiation of these two types of sentences suggests that they are indeed achieved by different combinatorial procedures, ordered by difficulty.

The novelty in three-word speech is usually thought of as the mastery of recursion or of hierarchical structure. While these are unquestionably key elements in children's novel skill, what emerges from this study is a new way of conceptualizing these, as well as the other components of the emerging syntactic knowledge. What children appear to learn in order to produce multiword sentences of any length are procedures for manipulating multi-element verbal information. The chief among these are procedures for the combination of information encoded in two different words, or for the splitting of information between two different words. In terms of a declarative linguistics, children need to learn the argument structure of individual predicate words and of rules for expressing pairwise predicate-argument relations in words. In procedural terms, children need to acquire the knowledge for splitting information between a predicate word and between the word providing the value for one of its arguments; as well as the techniques for unifying the information carried by the predicate and the argument words (e.g., Shieber 1986). Apparently, much of this learning is carried out at the one- and two-word stage, well before the child begins to produce three-word constructions.

The generation of verbal strings that consist of three words or more necessitates the construction of further procedures for information combination. In dependency-grammatical terms, a verbal string which is of three words long represents a discontinuous leap in the level of complexity of the construction: it cannot be produced on the basis of procedures for combining a predicate word and its argument but needs to involve some method for integrating two different syntagmatic relations. The present study highlighted the difference between linear and non-linear three-term structures; the former exploits the existing method of carrying the syntagmatic relation on the wave of the neighbourhood relation between two words, while the latter necessitates some procedure for disassociating neighbourhood from grammatical relatedness.

Although the difficulty presented by the processing of temporarily discontinuous dependency couples can be conceptualized as originating from an increase in short-term memory load, namely, as a general performance factor, it is probably better interpreted as a specific problem in dealing with temporarily open dependency relations. First, the memory load represented by an open dependency is minimal (one item of information stored for part of a second); a child in the middle of her second year should not find this amount of short-term memory storage prohibitive. Second, as we have seen, adults have an exceedingly limited capacity for processing open dependencies (Hudson 1992, 1993), far below the capacity of short-term memory; this suggests the workings of a different procedure than simply keeping some verbal information in temporary storage. The impression is that the difficulty of the young child in dealing with non-linear multiword constructions stems from the absence of a language-specific technique for creating syntagmatic connections between non-adjacent verbal items. If our preferred metaphor for the processing of relational information is that of unification between two partial information structures (e.g., Pollard & Sag 1987), it seems as if, at the onset of three-word speech, children may only be able to unify the information contained in adjacent words but do not know how to unify the information carried by non-adjacent ones (or to split information between them). As the type of information unification children master at a particular developmental stage determines the subset of the syntactic rule system they are able to construct, it is more appropriate to view limitation on these schemas as aspects of their syntactic competence proper than as language-external performance factors.

In order to test the generality of the findings of this study, a dependency analysis of other published corpora of early word-combinations (e.g., Bloom 1970, 1973, Bloom, Lightbown & Hood 1975, Bowerman 1973, Braine 1963, 1976, Brown 1973 and more) as well as the speech of a sample of children acquiring Hebrew is currently under way. Preliminary results suggest that Travis's pattern of initially producing linear three-word combinations may be a widely prevalent phenomenon in children acquiring English and Hebrew. Both of these are basically head-initial, SVO languages, in which it is possible to express the two simplest 3-term grammatical relations in linear patterns. Children apparently acquire these patterns first, and postpone the expression of the same grammatical relations in non-linear surface structures to a later stage in acquisition.

In summary, it is apparent that a dependency-syntax oriented acquisition model has considerable success in predicting what kind of sentences a child will first acquire on starting to combine more than two words in a single sentence. In making these predictions, the dependency-oriented developmental theory does not assume that a child has recourse to innate syntactic knowledge which is suddenly made available to the child. On the contrary, the acquisition of syntactic knowledge is seen as a continuous and gradual process, firmly built on the foundations of two-word speech and even of learning carried out in the single-word period.

Overall, the developmental evidence presented in this paper demonstrates that an acquisition theory adopting dependency grammar as its theoretical framework is successful in accounting for a series of developmental phenomena connected with the acquisition of syntactic knowledge. This finding, together with the limited success of phrase-structure oriented acquisition theories to account for the early stages of multiword speech, raises the possibility that a dependency grammar may be a better approximation to implicit grammatical knowledge in people than a phrase-structure type grammar. Consequently, the adoption of dependency theory as a linguistic framework of child language research may bring about a better understanding of the child's task in acquiring syntactic knowledge than we currently possess, exchanging such terms as innate higher-order syntactic constituents with units of knowledge more firmly anchored in observable developmental patterns.

REFERENCES

Abney, S. (1987). The English noun phrase in its sentential aspects. Unpublished doctoral dissertation, M.I.T., Cambridge, Mass.

Atkinson, M. (1985). How linguistic is the one word stage? In M. D. Barrett (ed.), *Children's Single-word Speech* (New York: Wiley).

Bloom, L. (1970). *Language Development: Form and Function in Emerging Grammars* (Cambridge, Mass.: M.I.T. Press).

Bloom, L. (1973). *One Word at a Time: The Use of Single Word Utterances Before Syntax* (The Hague: Mouton).
Bloom, L., Lightbown, P. & Hood, L. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development*, **40** (2, Serial No. 160).
Borer, H. & Wexler, K. (1987). The maturation of syntax. In T. Roeper & E. Williams (eds), *Parameter Setting* (Dordrecht: Reidel).
Bowerman, M. (1973). *Early Syntactic Development: A Cross-linguistic Study with Special Reference to Finnish* (Cambridge: Cambridge University Press).
—— (1976). Semantic factors in the acquisition of rules for word use and sentence construction. In D. Morehead & A. Morehead (eds), *Directions in Normal and Deficient Child Language* (Baltimore: University Park Press).
Braine, M. D. S. (1963). The ontogeny of English phrase structure: the first phase. *Language*, **39**, 1–14.
—— (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development*. 41, (1, Serial No. 164).
Branigan, G. (1979). Some reasons why successive single word utterances are not. *Journal of Child Language*, **6**, 411–421.
Bresnan, J. (1978). A realistic transformational grammar. In M. Halle, J. Bresnan & G.A. Miller (eds), *Linguistic Theory and Psychological Reality* (Cambridge, Mass.: M.I.T. Press).
Brown, R. (1973). *A First Language: The Early Stages* (Cambridge, Mass.: Harvard University Press).
Chomsky, C. (1969). *The Acquisition of Syntax in Children from 5 to 10* (Cambridge, Mass.: M.I.T. Press).
Chomsky, N. (1965). *Aspects of the Theory of Syntax* (Cambridge, Mass.: M.I.T. Press).
—— (1975). *Reflections on Language* (New York: Random House).
—— (1981). *Lectures on Government and Binding* (Dohrecht: Foris).
—— (1992). A minimalist program for linguistic theory. *MIT Occasional Papers in Linguistics*, **1**.
Dowty, D. R., Wall, R. E. & Peters, S. (1981). *Introduction to Montague Semantics* (Dordrecht: Reidel).
Elbers, L. (1990). The synchronic relation between two- and three-word sentences. Paper presented at the Fifth International Congress for the Study of Child Language, Budapest, Hungary.
Ewing, G. (1982). Word-order invariance and variability in five children's three-word utterances: a limited-scope analysis. In C. E. Johnson and C. L. Thew (eds), *Proceedings of the Second International Congress for the Study of Child Language*, Vol. 1 (Washington, DC: University Press of America).
Fonagy, I. (1972). A propos de la genèse de la phrase enfantine. Lingua, **30**, 31–71.
Gazdar, G., Klein, E., Pullum, G. & Sag, I. (1985). *Generalized Phrase Structure Grammar* (Oxford: Basil Blackwell).
Goldenberg, G. (1989). The contribution of semitic languages to linguistic thinking. *Journal of the Ancient Near Eastern Society Ex Oriente Lux*, **30**, 107–115.
Hawkins, J. (1983). *Word Order Universals* (New York: Academic Press).
—— (1990). A parsing theory of word order universals. *Linguistic Inquiry*, **21**, 223–261.
Hill, J. C. (1984). Combining two-term relations: evidence in support of flat structure. *Journal of Child Language*, **11**, 673–678.
Horgan, D. (1976). Linguistic knowledge at early Stage I: evidence from successive single word utterances. *Stanford Papers and Reports in Child Language Development*, **12**, 116–126.
Hudson, R. (1984). *Word Grammar* (Oxford: Basil Blackwell).
—— (1987). Zwicky on heads. *Journal of Linguistics*, **23**, 109–132.
—— (1990). *English Word Grammar* (Oxford: Basil Blackwell).

—— (1992). Dependency counts. Paper presented at a colloquium on 'Functional description of language', Charles University, Prague.

—— (1993). Do we have heads in our minds? In G. Corbett, S. McGlashan & N. Fraser (eds), *Heads in Grammatical Theory* (Cambridge: Cambridge University Press).

Hyams, N. M. (1987). The setting of the null subject parameter: a reanalysis. Boston University Conference on Language Development, Boston.

Ingram, D. (1979). Stages in the development of one-word utterances: transition to semantic relations. In P. L. French (ed.), *The Development of Meaning* (Hiroshima: Bunka Hyoron).

—— (1989). *First Language Acquisition* (Cambridge: Cambridge University Press).

Kazman, R. (1988). Null arguments and the acquisition of Case and Infl. Boston University Conference on Language Development, Boston.

McNeill, D. (1966). Developmental psycholinguistics. In F. Smith and G. Miller (eds), *The Genesis of Language: A Psycholinguistic Approach* (Cambridge, Mass.: M.I.T. Press).

Mel'cuk, I.A. (1979). *Studies in Dependency Syntax* (Ann Arbor: Karoma).

—— (1988). *Dependency Syntax: Theory and Practice* (Albany: State University of New York Press).

Menyuk, P. (1969). *Sentences Children Use* (Cambridge, Mass.: M.I.T. Press).

—— (1977). *Language and Maturation* (Cambridge, Mass.: M.I.T. Press).

Montague, R. (1974). *Formal Philosophy* (Edited by R. H. Thomason, London: Yale University Press).

Nelson, K. E., Carskaddon, G. & Bonvillian, J. D. (1973). Syntax acquisition: impact of experimental variation in adult verbal interaction with the child. *Child Development*, **44**, 497–504.

Nelson, K. E., Denninger, M. S., Bonvillian, J. D., Kaplan, B. J. & Baker, N. D. (1984). Maternal input adjustment and non-adjustments as related to children's linguistic advances and to language acquisition theories. In A. D. Pellegrini & T. Yawkes (eds), *The Development of Oral and Written Language in Social Contexts* (Norwood, NJ: Ablex).

Ninio, A. (1988). On formal grammatical categories in early child language. In Y. Levy, I. Schlesinger & M. D. S. Braine (eds), *Categories and Processes in Language Acquisition* (Hillsdale, NJ: Erlbaum).

Noonan, M. (1985). Complementation. In T. Shopen (ed.), *Language Typology and Syntactic Description. Vol II. Complex Constructions* (Cambridge: Cambridge University Press).

Peters, A. M. (1986). Early syntax. In P. Fletcher & M. Garman (eds), *Language Acquisition*, 2nd ed. (Cambridge: Cambridge University Press).

Pinker, S. (1984). *Language Learnability and Language Development* (Cambridge, Mass.: Harvard University Press).

Pollard, C. & Sag, I. (1987). *Information-based Syntax and Semantics. Vol I. Fundamentals* (Stanford: CSLI).

Radford, A. (1990). *Syntactic Theory and the Acquisition of English Syntax: The Nature of Early Child Grammars of English* (Oxford: Blackwell).

Robinson, J. J. (1970). Dependency structures and transformational rules. *Language*, **46**, 259–285.

Schlesinger, I.M. (1982). *Steps to Language* (Hillsdale, NJ: Erlbaum).

Scollon, R. (1976). *Conversations with a One Year Old* (Honolulu: University Press of Hawaii).

Shieber, S. M. (1986). *An Introduction to Unification-based Approaches to Grammar* (Stanford: CSLI).

Siewierska, A. (1988). *Word Order Rules* (London: Croom Helm).

Speas, M. (1990). *Phrase Structure in Natural Language* (Dordrecht: Kluwer).

Tomasello, M. (1992). *First Verbs: A Case Study of Early Grammatical Development* (Cambridge: Cambridge University Press).

Tutte, W. T. (1984). *Graph Theory* (Menlo Park, CA: Addison-Wesley).
Valian, V. (1986). *Syntactic categories in the speech of young children. Developmental Psychology*, **22**, 562–579.
Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, **40**, 21–81.
Zwicky, A. (1985). Heads. *Journal of Linguistics*, **21**, 1–30.

APPENDIX A

Travis's first 102 multiword utterances of three words or more, by their dependency structure and the order of their appearance

Dependency pattern Number*	Structure	Utterances
1	H (H D)	
	—> —>	Ride a bike; Throw da ball; Get the pencil; Top medicine that; Lock that Lulu; Hold da pencil; Hold da cigos; Open the umbrella; Hold da silk; Hold da jelly; Read this book; See Daddy's car; Close this window; Stuck this Weezer-pillow; See the paper; Ride this Mommy; Bye-bye Daddy's car; Catch the ball; Hold this ball; Step-in this pen; Bite this pigtails; Sit-down one pillow; See the rabbit; Bite this cracker; Hold this cream; Open this cracker; Bite the cracker; Break this bite; Hold the spoon; Find the stick; Hold this spoon; Move Daddy tray; Move the garbage; Big Daddy's tree; See the pizza; Hold the bubbles; Step-in that moo; Hit the wall; Stuck this Daddy; Open the button; Stuck that book; Open the top; Drop the checkbook; Step-in this doo-doo; Bite this ball; On the light; Move the chair; Look the Cinnamon; See the sheep; Read this color-book; Hold this poc; Play this silk; Drinking the bottle; Hold this blocks; Sit-down Weezer pillow; Stuck on bowl; Down on couch; Hold the grape-juice; Hold this paper; Play this crayon; Play this

Continued on next page

APPENDIX A – *continued*

Dependency pattern		
Number*	Structure	Utterances
		drum; Fall-down Maria's bike; Ride the pooh; Sit-down this bed
1A	H (H (H D)) —>—>—>	See the picture tiger
2	D_1 H D_2 <— —>	Cereal down rug; Tennis down rug; Chicken off hands; Mommy made book; Big-Bird ride horsie; Coffee down table; Linda made ice; Daddy made this; Peas off table; Nini over here; Weezer did it; Dana made this; Poker under car; Danny hit tennis; Daddy drive keys
2A	D_1 H (H_2 D) <— —> —>	Maria made this duck; Linda made that dress; Maria made that book; Mommy made this pictures; Dana made this balloon; Linda made this shirt; Dana made that color-book; Mommy made this picture; Mommy made that table
3	(D H) H <— <—	Big rock stuck; Creme sandwich here; Creme sandwich gone
4	H D_1 D_2 —> ———>	Stay here rug; Draw me man; Put-it on ring; Down this right-here; Stay here breakfast
4A	H (H_1 D) D_2 —> —> ————>	Read this book outside; Read this book again
5	D_1 D_2 H <———— <—	Doo-doo fork eat-it

Continued on next page

APPENDIX A – *continued*

Dependency pattern Number*	Structure	Utterances
6	H (D H) ———> <—	Weezer cat milk
7	(H D) H —> <———	Two rugs down

* Four-word constructions are numbered as the closest three-word pattern, with the addition of an 'A'.