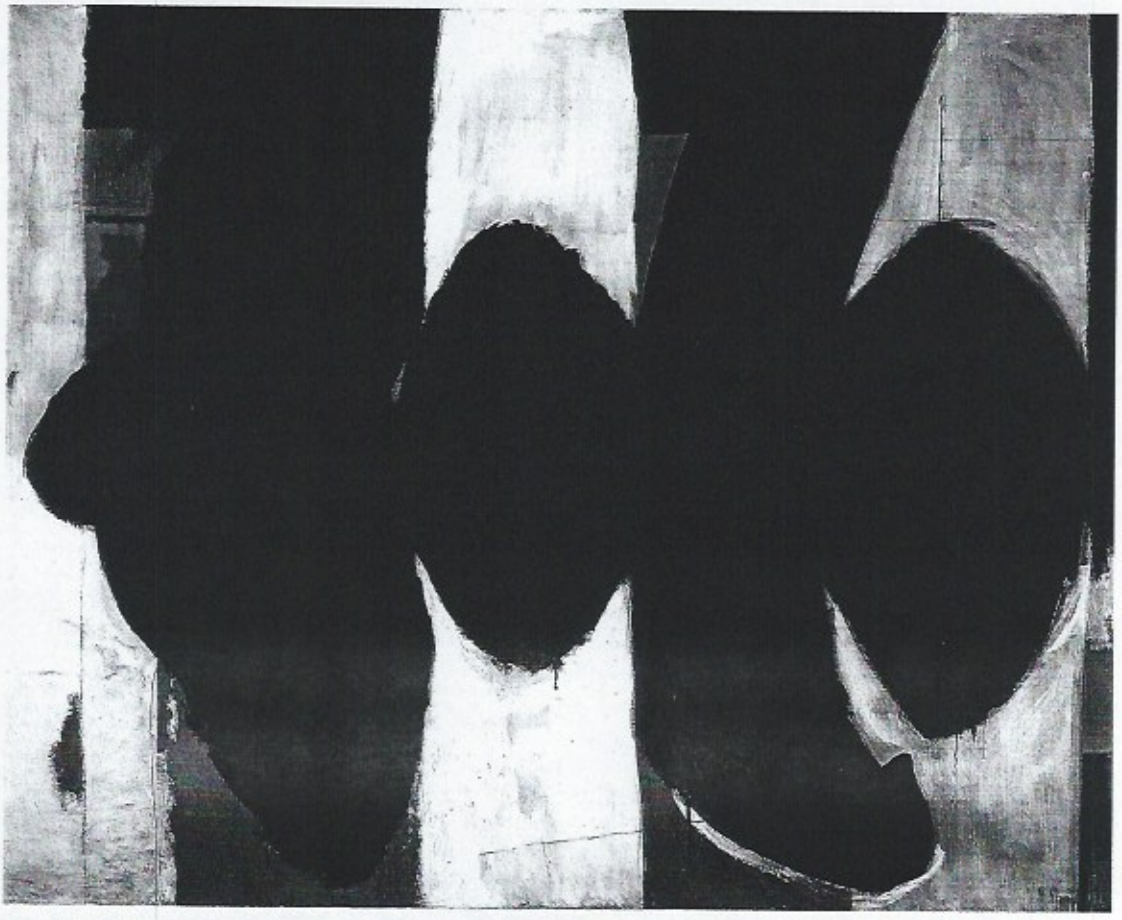
Linguistics

An Introduction to Linguistic Theory



Written by Victoria A. Fromkin (editor)

Susan Curtiss Bruce P. Hayes Nina Hyams Patricia A. Keating

Hilda Koopman Pamela Munro Dominique Sportiche

Edward P. Stabler Donca Steriade Tim Stowell Anna Szabolcsi

6

Acquisition of Word and Sentence Structure[1]

CHAPTER CONTENTS

6.0 Introduction

Now that we understand something of the morphology and syntax of adult grammars, we are in a position to look more closely at the question of how children acquire these aspects of grammar. As we have seen in the preceding chapters, grammars are very complex. Yet, each of us naturally acquires such a system without any special training and on the basis of unstructured and impoverished data. In this chapter we will examine morphological and syntactic development. We cannot cover all areas of child morphology and child syntax, but we will try to give a sense of what young children know about the structure of their language at various points in their development and how this is the same or different from adult linguistic competence.

6.1 Knowledge of Word Formation

Children begin to produce words sometime towards the end of the first year or beginning of the second year of life and learn words rapidly throughout the preschool years. But what do young children know about words and word formation? Chapter 2 discussed the structure of words and the variety and complexity of word structure. How and when do children acquire this knowledge?

Children appear to know the principles that govern word formation extremely early, suggesting that this kind of grammatical knowledge may be a part of Universal Grammar, the linguistic principles that are present at birth, discussed in Chapter 1. This knowledge about how words are constructed is attested in language development in a number of ways. We will illustrate this by discussing: (1) knowledge of what can be a word and (2) knowledge of morpheme-attachment, focusing on compounds.

Children typically produce novel words, that is, words that are not rote-learned formulas like *patacake* or *peekaboo*, early in the second year of life. The mean age for producing ten words is 15 months, and remarkably, even when these first words emerge, children appear to know the constraints and principles governing the ways morphemes may be combined to form a word. Certain kinds of morphemes may not emerge until much later, and errors of omission and mis-selection of morphemes may persist until the early school years, but violations of the basic constraints of morphological structure are not attested.

6.2 What Can be a Word?

Children's earliest speech consists of one-word utterances. This is often referred to as the **one-word stage**, or the **holophrastic** (which means expressing a complete sentence in one word) stage. From this point on, children appear to know not only what a word is, but what can constitute a word from the standpoint of the morphological structure of their language.

6.2.1 Language-particular constraints

How can we tell from the one-word utterances that very young children produce, that they know and are obeying both universal and language-particular morphological constraints? One kind of evidence is that children learning typologically different kinds of languages from the standpoint of morphology produce morphologically different kinds of words. From the outset of word production, for example, children demonstrate that they know whether a bare root of a major lexical category constitutes a well-formed word in their target language. In one-word speech, children

acquiring languages like English or Vietnamese, for example, where bare roots (e.g. *bottle, water, go, want, eat*) may be words, produce verbs and nouns without any bound morphology such as tense markers or plural markers attached. In contrast, children acquiring languages like Turkish, Italian, Greenlandic, or Hebrew, where a bare root without bound morphemes rarely, if ever, occurs as a well-formed word, do not produce bare roots which would be ungrammatical in their language. Children acquiring such languages produce inflected roots – roots with bound affixes attached – even at the earliest stages of language development.

6.2.2 Languages where a bare root is not a well-formed word

Let us examine acquisition in Turkish as an example.

(1) **Turkish** – a highly inflected language in which nouns take the form:

STEM–(PLURAL)–(POSSESSIVE)–CASE

and verbs take the form:

VERB STEM–MODE–TENSE/ASPECT–PERSON/NUMBER

The morphemes in parentheses are present when appropriate; the others are obligatory at all times. This appears to be a complex inflectional system which we might expect to take a long time to acquire. Yet, both noun and verb inflections are present at the one-word stage, and the entire set of inflections is mastered by or before 24 months of age. In fact, with respect to morphological structure, according to Aksu-Koč and Slobin, "... Turkish child speech is almost entirely free of error ... Most of the morphology – nominal and verbal – is used productively at the two-word period, before the age of 2."[2]

(a)–(c) (with the phonology somewhat simplified) are illustrative of early word productions in Turkish.

(The age of the child is represented in years; months. So 2;1 means 2 years, 1 month.)

(a)	bit-ti finish-past	'allgone'	(1;6)
(b)	kalem-i pencil-acc.	'the pencil'	(1;10)
(c)	sev-mi-eceg-im onu daha love-NEG-fut-1sg. 3sg:acc. more	'I won't love her anymore'	(2;0)

Note that the verbs and nouns are all inflected stems; in no instance are they bare roots. Other examples of early one-word utterances produced by children learning languages in which a bare root is not a grammatical (content) word are presented in (2)–(4) below, in each case attesting to the same developmental phenomenon: that children produce only words that are morphologically possible words in their language from the earliest stages of acquisition.

(2) **Hebrew** – a highly inflected language with unpronounceable, tri-consonantal (sometimes called triliteral) verb roots which appear as discontinuous morphemes, as discussed in Chapter 2. (Vowels, representing morphemes, are inserted into these roots.) These verb roots are obligatorily inflected for person, number, tense and gender. In the Hebrew examples, the capital letters represent the tri-consonantal roots and pluses (+) are used to show the morphemes contained in these complex roots.

a.	G+a+M+a+R -ti finish–past -1sg.	'I('m) finished'	(1;7)
b.	Ø +o+X+e+L -et eat-pres.-fem.sg.	'(I'm) eating'	(1;8)
c.	H+o+L+e+X -et walk-pres. -fem.sg.	'(Mommy) is walking'	(1;8)

(3) **Italian** – a language where verbs are inflected for person + number, and tense/aspect:

a.	pend-o io take-1sg. I	'I take'	(1;11)
b.	fai-te fai-te do-2sg. do-2sg.	'(You) do, (you) do'	(1;11)
c.	son-o giu be-1sg. downstairs	'(I) am downstairs'	(1;11)

(4) **Greenlandic** – a polysynthetic language with 318 inflectional affixes and over 400 derivational bound morphemes. A verb typically consists of a stem followed by derivational affixes, terminating in an obligatory inflectional affix; adjectives and other specifications on the noun are realized as inflections on the noun stem.

a.	nangia-ssa-ngil-anga be scared-fut.-not–1sg.indicative	'I shan't be scared'
b.	anar-tar-fi(k)ler-i-su-u-pput defecate-habit-place-agentive intran. particip.-be involved with-be-3pl.indicative sewage collector are	'They are the sewage collectors'

c. tuluar-suaq 'big raven'
raven-big
d. una-a-nngit-toq 'It's not that'
that-be-not-3sg.participial

6.2.3 Languages where a bare root is a well-formed word

Now contrast this acquisition pattern with that in which children are acquiring languages that do allow monomorphemic words, and where monomorphemic words are the basic word type. Such languages include English and Vietnamese. By examining the early words produced by children learning such languages, we find that typical early words include words like *look, see, want, bottle, cookie; water, rice, father, chopsticks, eat, finish, come* – all bare roots, and early two- and three-word utterances are also typically formed with bare-root words, as illustrated in (5)–(6):

(5) English
a. Want bottle
b. No sleep
c. Want more cookie
d. See Mommy car
e. Look my shoe

(6) Vietnamese
a. an dua 'eat with chopsticks'
eat chopsticks
b. ve nuoc 'finished with the water'
finish water
c. bo den 'father is coming'
father come
d. bo an com 'father is eating rice'
father eat rice
e. me roi com 'mother is cooking rice'
mother cook rice

A well-formedness constraint on the structure of content words (Nouns, Verbs and Adjectives in particular) in a given language determines whether a bare root can be a word. A revealing example of how strongly children's language is guided by this constraint in early stages of acquisition can be found in a surprising fact about the acquisition of verb forms in Turkish. In Tukish, the second-person imperative emerges late relative to most other forms of the verb. This is surprising because the second-person imperative (e.g. '(You) Hold me') is typically among the two earliest verb forms to be acquired cross-linguistically, and in Turkish it is just the stem and bears no

affixes. Why would children learning Turkish acquire this morphologically simple verb form, a form they hear very frequently in speech directed to them, later than more complex forms of verbs? One explanation that accounts for the patterns of acquisition we have seen is that the early grammar of the child acquiring Turkish has the constraint that well-formed verbs may not be bare roots in Turkish, but the second-person imperative bears no affixes; it is the same as the stem. The second-person imperative,

Acquisition of signs SIDEBAR 6.1

The sign languages used by the deaf throughout the world show similar constraints on word formation similar to those of spoken languages. In sign languages, morphemes consist of specific gestures or combinations of gestures performed in space. Some sign words consist of only one morpheme, some signs consist of more than one morpheme. Each sign language has its own morphemes and its own rules about what can be a well-formed sign/word. Children learning sign languages use the same word-structure principles in forming their first words as children acquiring spoken languages. This indicates that children are predisposed to look at the grammar of the language they are learning in just the right way to determine what the basic word type of their language is – whether it is a sign language or a spoken language. In American Sign Language (ASL), for example, there is rich verbal morphology. In Chapter 2, we saw that SLQ Zapotec verbs have separate forms for six different verb aspects (somewhat like tenses). Similarly, in ASL, some verb stems have different forms for more than fifteen distinct verb aspects and also carry subject and object agreement markers on the verb stem. Therefore, (a) and (b) are signed differently, as are (c)–(g).

(a) You give to me.
(b) I give to you.
(c) Olivia talks (on one occasion).
(d) Olivia talks all the time.
(e) Olivia talks over and over again.
(f) Olivia talks for hours/endlessly.
(g) Olivia talks over a long period of time.

However, these aspectual and agreement morphemes are obligatory for only a subset of high-frequency verbs in the language. Children acquiring ASL treat their target language as a bare root-type language like English because, for the most part, it is. Thus they produce uninflected roots as their first words, even for verbs that obligatorily take verbal inflections. Moreover, because this subset represents a set of exceptions to the basic word-formation pattern of the language and not the rule, children learning ASL begin to use some of these inflections only between $2\frac{1}{2}$ and 3 years of age; and acquisition of the full set of inflectional morphemes follows a rather long course of development.

therefore, violates this constraint, and emerges only later, when the child's grammar is ready to handle such exceptions.

Turkish and English illustrate a linguistic contrast in what can be a morphologically well-formed word; but in each case, children obey the morphological well-formedness constraint on words in their language in their acquisition of word formation rules.

6.3 Morpheme Attachment

We have seen that children know whether bare stems can be words in their language, but when do children know the rules and constraints for putting morphemes together in creating polymorphemic words? Even at the earliest stages of language development children do not seem to make word-internal morpheme attachment errors, even when they are constructing words involving many morphemes, such as verbs in Turkish, Greenlandic, or Quiché. That is, they never use a prefix as an infix or suffix (for example, children acquiring English do not say *@ingplay* instead of *playing*, and children acquiring French do not say *@uispart* instead of *parti(s)*); they do not put the morphemes in the wrong order, putting the head of a compound first instead of last (*@brushtooth* instead of *toothbrush*), nor do they attach suffixes in the wrong order, even when there are more than one (@catcheser instead of *catchers*). (Note: @ indicates a form or sentence that is unattested.)

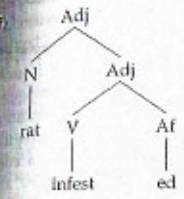
The fact that children always attach morphemes in the grammatically correct order, including the placing of inflectional suffixes after derivational morphemes shows that children from a very early age have an understanding of the hierarchical structure of words and the distribution of different types of morphemes discussed in previous chapters.

6.3.1 Compound formation and Level Ordering

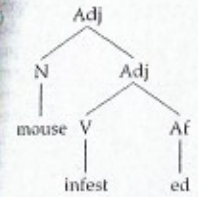
Young children also show their knowledge of the rules for the formation of compounds, such as those discussed in chapter 2. For example, in forming compounds, one needs to be sensitive to the distinctions between nouns that take a regular plural and those that don't. A noun that takes a regular plural, that is, that forms its plural by rule, may not enter a compound in its plural form because the pluralization rule follows the process of compound formation (for example, *rat-infested* but not **rats-infested*), but a noun whose plural form is irregular, that is, whose plural is not formed by rule, may optionally be used in a compound in its plural form (both *mouse-infested* and *mice-infested* are acceptable). Irregular plurals may appear inside compounds because the irregular plural form is stored along with the singular form in the lexical entry for that noun, and so may enter the

compound as part of the noun stem. Compare (7), which illustrates the derivation of the word 'rat-infested', where only the singular form of the noun may enter the compound, with (8) and (9), which illustrate the derivations of the words 'mouse-infested' and 'mice-infested', respectively, recalling that either 'mouse' or 'mice' can be used in the compound because the irregular plural form, 'mice', is stored along with the singular form, 'mouse', in the same lexical entry:

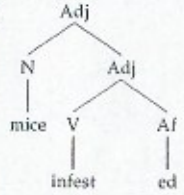
(7)



(8)



(9)



6.3.2 The Level Ordering Model

A model by which irregular inflectional (and idiosyncratic derivational) affixes are attached before regular derivational and inflectional rules apply is referred to as the 'Level Ordering Model,' because the model hypothesizes distinct, sequential levels of operation by which complex morphological forms are constructed, with specific processes occurring at one and

only one level, not before, not later. (10) presents a description of the kinds of morphemes which may be attached at each level:

(10) **Level Ordering Model**
- Level 1 – unproductive or semantically irregular derivational or inflectional affixes
- Level 2 – productive and semantically regular derivational affixes
- Level 3 – regular inflectional affixes

Recall that in this model of morpheme attachment, no level may apply before a preceding level. Thus, Level Ordering predicts that in forming plural compounds, no regular plurals may occur inside the compound, because regular plural markers are attached at level 3, after all derivational and irregular inflectional affixes have been attached. However, irregular plurals may be attached at level 1, and therefore appear inside a compound. (11) illustrates how the compounds *rat-infested*, *mouse-infested*, and *mice-infested* would be formed under Level Ordering, and (12) illustrates how the plural compounds *rat-swallowers*, *mouse-swallowers* and *mice-swallowers* would be formed under Level Ordering.

(11)	(a)	(adult-preferred)	rat infest	mouse infest
		Level 1	——	——
		Level 2	rat infest-ed	mouse infest-ed
		Level 3	——	——
		Output	rat infested	mouse infested
	(b)	adult dispreferred	rat infest	mouse infest
		Level 1	——	mice infest
		Level 2	rat infest-ed	mice infest-ed
		Level 3	——	——
		Output	rat infested	mice infested
(12)	(a)	(adult-preferred)	rat swallow	mouse swallow
		Level 1	——	——
		Level 2	rat swallow-er	mouse swallow-er
		Level 3	rat swallow-er-s	mouse swallow-er-s
		Output	rat swallowers	mouse swallowers
	(b)	adult dispreferred	rat swallow	mouse swallow
		Level 1	——	mice swallow
		Level 2	rat swallow-er	mice swallow-er
		Level 3	rat swallow-er-s	mice swallow-er-s
		Output	rat swallowers	mice swallowers

We can see, then, that in forming plural compounds, it is never possible to pluralize the first element in a plural compound if it takes a regular plural inflection, because regular plural inflection is carried out at level 3, and therefore may be attached as a suffix only to the entire compound. However, it is possible to pluralize the first element in a compound if it takes an irregular plural, because irregular pluralization occurs at level 1, before any other affixation occurs. Pluralizing the first constituent of the compound in such cases is optional, however, and highly dispreferred among adults. Adults rarely, if ever, use this option to construct plural compounds like *mice-eater* or *teeth-swallower*. Some researchers have hypothesized that Level Ordering is part of UG and thus need not be learned.

6.3.3 An experimental test

How can one test the Level Ordering hypothesis and the hypothesis that Level Ordering is part of UG and so constrains children's word formation, without having to be learned? Peter Gordon (1985) tested these hypotheses by studying whether children obeyed the principles of Level Ordering in forming plural compounds. If children adhered to the constraints of the model, they would never use a noun which takes a regular plural inside a compound, but would optionally allow a noun which has an irregular plural form to appear inside a compound. If, further, the youngest children as well as the older children followed the model, then it would suggest that Level Ordering was not something they had to learn over time, but could be part of UG.

Gordon tested 3- to 5-year-old children's knowledge of how to form plural compounds, such as *dog-catchers*. For the first noun of each compound he used both kinds of nouns: words like *rat*, *bead*, and *hand*, which take regular plurals (e.g. *rats*, *beads*, *hands*), and words that have irregular plural forms (e.g. *tooth*, *foot*, *mouse* and *man*). He showed the children objects and using Cookie Monster to ask them questions, elicited a plural form for the object (for example, *'hands'*) and immediately following the production of that plural, elicited a compound using that noun. To elicit the compound, he had Cookie Monster ask, "What do you call someone who eats X" (e.g. hands)? So, for example, Cookie Monster would ask, "What do you call someone who eats hands?" The compound that the child produced with each noun indicated whether or not the child's grammar permitted that noun to occur inside of compounds in its plural form. Recall that adult English speakers strongly prefer to use the singular form of words inside compounds, even when the grammar allows plural forms to occur; thus, children rarely if ever hear plural nouns inside of compounds.

There were three important results in Gordon's experiment. First, by 3 years-of-age the children knew how to form plural compounds, such as *rat-eaters* or *mice-swallowers*. Second, the children consistently followed the

Level Ordering Model. They reduced the regularly pluralized nouns to singular forms inside the compounds, that is, they formed [Noun sg. – Vers] compounds for nouns taking regular plurals (e.g. *rat-catchers*) and created [N pl. – Vers] compounds with nouns forming irregular plurals (e.g. *mice-catchers*, *teeth-swallowers*). These results indicated that in forming complex words, their grammars already followed level ordering constraints. Children demonstrated this even with their errors, for children who produced incorrect regular plurals such as *mouses* or *tooths*, never produced compounds such as **mouses-eater*. Third, those children who knew the correct irregular plural form of the nouns used in the experiment formed compounds sometimes with the singular (*tooth-swallowers*) and sometimes with the irregular plural, producing forms such as *teeth-swallowers*. Since an examination of speech spoken to children indicates that they rarely if ever hear plural compounds in the input, and probably never hear forms such as *teeth-swallowers* or *mice-catchers*, forms highly dispreferred by adults, their knowledge of how to form plural compounds is an instance of the acquisition of linguistic principles or rules despite the 'poverty of the stimulus' discussed in chapter 1, lending support to the hypothesis that this knowledge is unlearned and part of Universal Grammar. Gordon's work is also an illustration of how acquisition facts can inform or support hypotheses developed solely to account for the adult grammar.

Gordon's results are not unusual. By the time children know how to form and produce compounds, typically by 3 years-of-age, they appear to know almost everything there is to know about forming compounds, such as those discussed in chapter 2. This includes where to place the head constituent and therefore (by and large) how to interpret the compound, what morpheme attachment must precede or follow the formation of the compound (e.g. inflectional morphemes follow derivational morphemes – Level Ordering, again), and how a compound's stress pattern differs from phrasal stress in English. Errors by children in using phrasal versus compound stress appropriately are rare, even when a compound and a phrase consist of the same words as in (13) and (14), where the stressed syllable is in upper case:

(13) a blueBIRD (a bird that is blue)
(14) a BLUEbird (a particular species of bird)

6.4 Word-Formation Errors

Children's production of morphologically complex forms is not error-free, but their knowledge of the language's word-formation rules is reflected in the errors they do make. Such errors include creating words using productive word-formation rules of their grammar, words such as *sweeper* or *typer*, for example, which are later replaced by existing words in the language that express the same meanings – *broom* and *typist*, in this case.

Most word-formation errors are rare, but not everything about word formation develops early. Children can take a long time to learn the particulars of word formation which are idiosyncratic and for which there are no general rules or principles that produce them. In English, for example, it is necessary to learn by rote which (non-productive) derivational affixes go with which roots and what they mean. Children make frequent errors of this type along the way, as illustrated by (15) and (16), errors made by RG at the age of 2;6 and 4;2, respectively:

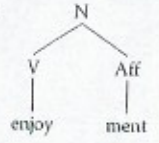
(15) (Coming in from playing outdoors) (2;6)
R: I want a yellow bicycle – no, some 'sickle.'
(Still trying to retrieve the word, R pointed to the freezer, and said)
Popsicle!

(16) (R's blanket was in the dryer and R was anxious to have it.) (4;2)
R: I think it's had enough **drymeat*.

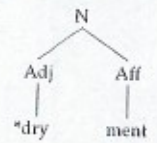
In (15) R confused the morphemes *-sicle*, which occurs in words like *hedgesicle* and *icicle*, with *-cycle*, as in *bicycle* and *tricycle*. Both morphemes form nouns, and both are pronounced the same, but they have different meanings and attach to different morphemes. They are homophonous morphemes – same sounds, different meanings, similar to word homophones like *to*, *too*, and *two*. However, in R's lexicon they were still represented as the same morpheme.

In (16), R attached the wrong derivational morpheme to the adjective *dry* to form the noun she wanted. (She should have used *-ing*.) But note that she selected a derivational morpheme which *does* form nouns, but forms them from verbs, not adjectives, as shown in (17a) and (17b):

(17) a.



b.



Except for errors in the use of inflectional morphology (which will be discussed below), the word-formation errors children make involve forms which must be memorized, and/or idiosyncratic properties of words which must be listed in the lexicon rather than formed by general principles of morphology. The long period over which these forms are mastered and the many bumps along the way contrasts sharply with the relative rapidity

and ease with which the bulk of the grammar is attained, pointing to the difference between what has been viewed as biologically-programmed maturation or unfolding of grammatical knowledge in development, and language *learning*.

6.5 A Note About Function Words

Although early speech of children consists largely of lexical or open class words, some closed class or function words also emerge early, frequently at the onset of two-word speech. Research indicates that children appear sensitive to the function-word/lexical content-word distinction in both comprehension and production, early in development. Before mastering their phonological form, children appear to use simple (often a single vowel) phonetically neutral forms, referred to by some researchers as 'monosyllabic place holders' (MPHs) for function words. These MPHs appear where function words should occur, and knowledge of specific categories of function words is revealed through their appropriate distribution relative to lexical content words – articles (e.g. *a, the*) before nouns in English and Italian, for example. Through an examination of the use of such MPH morphemes in combination with content words, then, it has been found that, in some cases, before children have acquired complete knowledge of the syntax or pronunciation of certain morphemes, they appear to know what kinds of morphemes they are (lexical or functional), and the rules governing their distribution relative to other morphemes.

6.6 The Syntax of Child Grammars

In this section we will focus on syntax development. We cannot hope to cover all areas of child syntax, but we will try to give a sense of what young children know about the syntactic structure of their language at various points in their development and how this is the same and yet different from adult linguistic competence.

Sometime around age 2 children begin to produce multi-word utterances. Initially, their utterances are only two or three words long and the form of these early sentences is different in various respects from the adult 'target' language. For example, the following utterances are typical of the 2-year-old English speaking child:

(18) a. Make a choo-choo train.
b. Mommy throw it away.
c. No want this.
d. Not making muffins.

e. Man sit down.
f. See my doggie?
g. What cowboy doing?

It is important to note, however, that despite its apparent deviance, the child's language during this early stage (and any other developmental stage) is remarkably regular, and to the extent that it is different from the adult language, these differences are quite systematic. For example, we see in the sentences in (18) that children often omit subjects. However, English-speaking children typically do not leave out objects. We also see that at this age children may omit tense/agreement inflection from the verb, as in (18). Uttered by an adult this sentence would be *Mommy threw it away* or *Mommy throws it away*. On the other hand, progressive *-ing* occurs quite regularly during this stage, as illustrated in (18d, g).

Most of us find child language charming and are amused by the 'errors' that children make. But the child's language is not simply a haphazard approximation to the target. Rather, at each stage of development the child has an internally consistent, rule-governed cognitive system – a grammar, with specific properties that can be studied (just like the grammars of adult speakers). As discussed in chapter 1, the child begins with a universal grammar – UG – template, which specifies the general form of grammar, and then fill in the language-specific details – the gaps in the system – through experience with the particular language in his environment; some gaps are filled in quickly, others take more time, and there may be misformulations along the way, giving rise to acquisition stages. We may think about grammar development as involving the construction of a successive series of grammars. This is schematically shown in (19), where each stage of development is a grammar (G_n), beginning with the principles and structures made available by UG (G_0) and terminating in the adult grammar of a particular language (G_a).

(19) $G_0, G_1, \ldots, G_n, \ldots, G_a$

One of the interesting features of language development is the uniformity that we observe across children. Children raised under very different circumstances and who may otherwise differ from one another in all sorts of ways go through very similar stages in their development of language. For example, the **telegraphic stage** typified by utterances such as those in (18), in which certain elements such as articles and auxiliaries are missing, is common to all children (with some variation depending on the particular target). Even children acquiring very different languages, for example signed and spoken languages, show some striking parallels in their language, as discussed above. These commonalities that we find provide further evidence of a rich innate component to language. Otherwise, why