DE
—
G

DE GRUYTER
OPEN

ANAT NINIO
The Hebrew University of Jerusalem

VARIABLES AND VALUES IN CHILDREN'S EARLY WORD-COMBINATIONS

A model of syntactic development proposes that children's very first word-combinations are already generated via productive rules that express in syntactic form the relation between a predicate word and its semantic argument. An alternative hypothesis is that they learn frozen chunks. In Study 1 we analyzed a large sample of young children's early two-word sentences comprising of verbs with direct objects. A majority of objects were generated by pronouns but a third of children's sentences used bare common nouns as objects. We checked parents' two-word long sentences of verbs with objects and found almost no bare common nouns. Children cannot have copied sentences with bare noun objects from parents' two-word long sentences as frozen chunks. In Study 2 we raised the possibility that children's early sentences with bare nouns are rote-learned 'telegraphic speech', acquired as unanalyzed frozen chunks from longer input sentences due to perceptual problem to hear the unstressed determiners. To test this explanation, we tested the children's speech corpus for evidence that they avoid determiners in their word-combinations. The results showed that they do not; in fact they generate very many determiner-common noun combinations as two-word utterances. The findings suggest that children produce their early word-combinations of the core-grammar type by a productive rule that maps the predicate-argument relations of verbs and their semantic arguments to head-dependent syntax, and not as frozen word-combinations. Children mostly learn to use indexical expressions such as pronouns to express the variable semantic arguments of verbs as context dependent; they also employ bare common nouns to express specific values of the arguments. The earliest word-combinations demonstrate that children understand that syntax is built on the predicate-argument relations of words and use this insight to produce their early sentences.

Key words: syntactic development, predicate argument, frozen chunks, determiners, parental input

Introduction

In a recent study it was found that parents' two-word utterances express-ing the three core syntactic relations, namely, the subject-verb, verb-object

Address for correspondence: Anat Ninio, Department of Psychology, The Hebrew University, Jerusalem 91905, Israel. E-mail: anat.ninio@huji.ac.il

and verb-indirect object relations, express the verbs' complement in a very large proportion of cases with the help of pronouns and determiners, and some proper nouns (Ninio, 2014). Of the total 7,896 tokens of two-word sentences expressing core grammar combinations in parental speech, 6,401 or 81.1% of the complements to the verb were pronouns; 4.0% were proper nouns, while common nouns occurred in just 4.1% of sentences. These results are not surprising as two-word sentences allow only a single word to serve as the complement of the verb, and single-word referential expressions will tend to be pronouns and proper names, rather than common nouns. The reason is that common nouns referring to specific objects usually require a determiner in English, hence cannot be single-word expressions; single-word occurrences are reserved for unusual usages such as a plural noun referring to a type of objects and not to specific objects (e.g., in *'Like bananas?'*).

The special advantage of pronouns and proper nouns in parental two-word utterances expressing core grammar is that they make the relative role of the two words – the verb and the nominal – transparent. Pronouns and proper nouns are very special type of words in that the only way they can function in sentences is to refer to some entity (Reboul, 2001). When they accompany a word with the semantics of a verb, always encoding a complex event such as an action, it is obvious that the referential word must indicate some entity that serves as the semantic associate of the event depicted, such as its actor or object. Assuming that the child already knows the vocabulary items and can guess the meaning of the sentence from the non-linguistic context (Macnamara, 1972), the respective roles of the two words can easily be identified.

The two-word sentence containing a verb and a pronoun appears as the combination of an act of reference to an object and an act of predicating for it an argument role in the event-description given by the verb. For instance, in the two word sentence *'Push that'* which is a request to push some indicated object, the requested action of pushing is encoded by the verb *push,* and the object on which the action of taking is to be performed is coded by the pointer-word *that.* As the semantic-logical role of the object is dependent on the meaning of the verb, the relation can only be conceptualized as that of asymmetrical dependency.

The model of syntactic development proposed by Ninio (2014) is that children learn how to build multiword utterances from parental two-word utterances of this kind. The advantage of such short input sentences is that they contain transparent information on the binary dependency/merge relation responsible for syntactic connectivity. Children can learn from such utterances that word-combinations are built with the help of the asymmetrical Merge (Chomsky, 1995) or Dependency (Tesnière, 1959) operation connecting syntactic Heads and Dependents. In fact, parental two-word sentences are syntactic atoms. In dependency terminology, an atom of syntactic structure consists of two words, one of which

is a *head* (*governor*) of the other, its *dependent*. The head-word determines the occurrence in the sentence, positioning, morphological form, and semantic role of the dependent word. Semantically, the dependent's role is to modify the lexical meaning of the head. Syntactic relations encode the logical-semantic relation between the two words so that one of the words is a predicate and the other, its argument. In a complement relation (such as verb – direct object) the predicate is the head and the argument, its dependent; in an adjunct relation (such as attributive adjective – noun) the predicate is the dependent of the argument. As predicates are logical-semantic functions on variable arguments and are similar to mathematical functions such as $f(x)$ or $f(x,y),$ for a given predicate word, its logical argument(s) can take any number of different values in an actual sentence, hence the same predicate-specific coding rule expressing the predicate-argument relation generates an infinite number of different word-combinations.

The fact that the syntactic atoms modelled in the two-word parental input use mostly pronouns and proper names, that is, indexical referential terms whose meaning varies with each different context of use (Peirce, 1865/1982), signify that these sentences contain natural variables for dependents. If children indeed decipher the asymmetrical dependency structure – semantic and syntactic – of the sentences, as the model of development hypothesizes, they should be able to use the same combinatory rules to generate an infinite number of novel sentences of their own.

We already know that in their early word-combinations, young children indeed use the same verbs as the parents in the great majority of cases. In 96.0% all spontaneous tokens of the core grammatical relations produced by children in the 2;0-2;6 age range, the verb used was one that occurred in parental two word utterances expressing the same core grammatical relation (Ninio, 2014). Given that there is a wide consensus, based on present day linguistic theory as well as developmental data, that syntactic combinatory rules are lexical-specific rather than abstract, and thus require the learning of each verb's combinatory possibilities, these results support the learning model according to which children base their early word-combinations on parental two-word sentences. However, these results do not answer the crucial question, what exactly do children learn from these parental models? Since parental two-word sentences present transparently asymmetrical predicate-argument couplets expressed in clear surface patterns, Ninio proposed that children immediately learn from them to generate productive word-combinations of their own. Moreover, as parent's sentences contain natural variables such as pronouns as complements, the patterns learned are predicate-argument structures expressed as syntactic functions with a variable complement for the argument element. This proposal agrees well with previous suggestions by several investigators that the earliest two-word combinations (including the so-called pivot-open combinations) are productive patterns consisting of predicates and their logical arguments (e.g., Gentner, 1978; MacWhinney, 1982;

Ninio, 1988; Braine, 1976; Bloom, 1970; Bloom & Lahey, 1978; Deuchar & Quay, 2000; McCune-Nicolich, 1981; McCune & Vihman, 1999).

This is, however, not the only possibility for characterizing children's earliest word-combinations. In particular, a well-regarded model of syntactic development proposes that young children's first set of multiword utterances are learned as unanalyzed, stored chunks, from which productive combinatory rules are abstracted only at a later stage (Arnon, 2010; Diessel, 2013; Ellis, 2003; Lieven, Pine, & Baldwin, 1997; Lieven, Behrens, Speares, & Tomasello, 2003; Lieven, Salomo, & Tomasello, 2009; Pine & Lieven, 1993). This model of development also can account for the finding mentioned above, that is, that children's core-grammatical sentences almost exclusively use the same verbs as parents' two-word utterances with the same syntactic relations. Unless proven otherwise, it is thus possible that very young children use the same verbs as parents do when generating sentences expressing the subject-verb or verb-object relations because they learn parental sentences as unsegmented chunks and, when generating their own, they reuse the parental sentences verbatim.

This alternative model of development deserves a more detailed presentation. A recent handbook chapter by Arnon (2011) summarizes the approach:

"Usage-based and exemplar models suggest that children learn language by abstracting over stored utterances – they start out with largely unanalyzed chunks that are then analyzed and segmented (Abbott-Smith & Tomasello, 2006; Bod, 1998, 2006; Tomasello, 2003). [...] Young children make use of 'frozen' chunks in their early productions (e.g., Lieven et al., 2003), while older ones (three-year-olds) remain sensitive to the properties of four-word phrases (Bannard & Matthews, 2008; Matthews & Bannard, 2010). Taken together, these findings support the idea that multi-word chunks are part of children's initial linguistic inventory. Extending these ideas to construction learning suggests that children's early constructions may consist of frozen chunks, which are then analyzed and segmented to create schemas and slots (Bod, 2006; Lieven et al. 1997; Lieven et al., 2003, 2009)." (pp. 99-100)

In another recent handbook chapter, Diessel (2013) says: "The earliest utterances that children produce consist of isolated words and holophrases, i.e. unanalyzed chunks of multi-word expressions that are learned as frozen units." (pp. 351-352).

The first appearances of the frozen-units concept was in publications from the 1990s, and it stemmed from the researchers' caution not to over-estimate children's early syntactic capacity by treating any and all ostensibly multiword combinations as representing freely produced phrases created by children's syntactic rule system. For instance, the first combination with some predicate such as a verb not previously used by the child was coded as frozen, even if within a

very short time (even minutes) the child produced a different combination with the same verb. The same methodological caution still informs the claims made by present-day researchers that children begin the production of multiword utterances by using frozen chunks. For example, Lieven et al. (2009) define between 25% to 50% of children's early multiword productions as frozen utterances, based on the absence of evidence to the contrary. This coding decision is a methodological decision, and the researchers warn that it is not by itself evidence that children indeed begin multiword speech with frozen chunks.

However, in the meantime a new theoretical factor has joined the methodological caution not to over-estimate children's knowledge, and that is the claim that all productive syntactic rules are to be derived from large sets of previously unanalyzed multiword sentences by some process of distributional learning. For instance, in a recent publication (Bannard, Lieven, & Tomasello, 2009) a statistical procedure is described by which slot-and-frame formulae are derived from a large number of unanalyzed sentences stored by the learner's language system. As the cite from the handbook chapters makes clear, if productive formulae of any kind are thought to be derived from stored unanalyzed input sentences by a process of statistical or distributional analysis, prior to the extraction of such formulae children use in their speech chunks of frozen multiword input utterances stored in their lexicon as if they were single-word units.

It should be noted that the option that a large proportion of children's early two-word sentences are frozen units learned verbatim from the input is different from pointing out that a few exceptional early utterances are. It is undoubtedly true that when an English-speaking child barely produces any multiword speech, her saying something that sounds like *"Whereda bottle"* is unlikely to be the result of the productive generation of the four-word sentence *"Where is the bottle"*, and it is more likely that the expression *"whereda"* is a frozen, un-segmented version of the three-word phrase *"where is the"*. Peters (e.g., 1983) presented many such early segmentation errors in her publications. MacWhinney (1982) summarizes the 14 types of evidence used by researchers for word-strings being rote-learned and unanalyzed, including the criterion of being a 'precocious string' that have been used in discussing the example above. However, MacWhinney also pointed out that such rote-learned units are a small minority in children's early speech, saying "we have almost no evidence for rote units that extend beyond the level of the kinds of prosodic units that can be perceived as if they are words" (p. 88).

In the present study, we shall test the two alternative hypotheses by closely examining children's earliest and shortest word-combinations – the two word long ones. Some reports of children's very first word-combinations, for instance by Bowerman (1976, p.157), suggest that that the character of children's early multiword utterances supports the hypothesis of productivity and does not suggest that the earliest sentences were taken over as frozen

word-combinations from parental speech. Bowerman's daughter Eva began to produce word combinations at about 17 1/2 months, and right away in the first week she produced a large number of constructions involving the word *"want"*, in sentences like *'want bottle', 'want juice', 'want see',* and *'want change'.* The variety of noun complements suggests that Eva mastered a productive rule by which *want* can be combined with any word that specifies the thing or activity wanted (Bowerman, 1978, pp. 378-379). In addition, none of Eva's early two-word combinations with the verb *want* and its semantic object could have been taken over as verbatim copying of a frozen expression from parental speech as parents do not drop the determiners and other functional words missing from Eva's combinations.

Another child whose longitudinal acquisition data is available is Travis (Tomasello, 1992), whose acquisition of transitive constructions were analyzed in Ninio (1999). Travis acquired a large number of new verbs in the verb-object (VO) pattern in a very short time, so that within two months she produced 30 different verbs in this pattern. Among Travis's first sentences with each of her first 30 different verbs getting a post-verbal direct object, there were some pro-nouns, demonstratives and proper nouns serving as the direct objects, but in no less that 19 cases (63.3%), in the very first sentence expressing some verb with a direct object, the expression was a bare common-noun. For example, Travis said *"bite finger", "bring chair", "catch rocks", "driving car", "hit ball"* and *"touch light".* As in the case of Eva, such early word combinations with verbs and their semantic object could not have been copied verbatim from parental speech; instead, Travis's results suggest that her early sentences were already generated with variable syntactic roles that can be freely filled with contextually appropriate referential terms.

Although the longitudinal results are suggestive, the two alternative models of development should be tested on a large sample. In the following, we shall test the alternative developmental hypotheses on a large English-language sample. We shall examine the production of sentences ostensibly expressing the verb-direct object (VO) syntactic relation in children's speech. We shall focus on children's earliest and shortest sentences, namely, two words long VO sentences. We will examine the distribution of different kinds of expressions serving as direct objects in order to see if the data support either of the two hypotheses. The hypothesis of children learning initially chunks of multiword utterances (frozen or rote-learned) will be tested against the alternative hypothesis that children's earliest utterances are generated with productive rules with variable syntactic roles that can be freely filled with any contextually appropriate referential term.

We shall concentrate on the phenomenon well exemplified by Eva's and Travis's early VO sentences, that is, the omission of determiners and other closed-class elements and the use of bare common nouns as nominal phrases. That is, the major type of evidence we are considering is 'telegraphic speech'.

Telegraphic speech

The use of bare nouns instead of a determiner-noun phrase as a referential expression is thought to be highly characteristic of English-speaking children's two-word speech. The term 'telegraphic speech' for this kind of word-combinations was coined by Brown and Fraser (1963). Brown and Fraser, as well as Brown and Bellugi (1964), Ervin-Tripp (1966) and others pointed out that children's early multiword utterances tend to omit closed-class words such as articles, auxiliary verbs, copulas, prepositions and conjunctions, compared to the sentences adults typically say in the same circumstances. Children's sentences tend to include mostly open-class or substantive words such as nouns, verbs, and adjectives. Although it is not yet known how general the phenomenon is, in some texts the earliest stage of word-combinations is called the stage of telegraphic speech.

The reasons for children's producing 'telegraphic speech' are not yet clear. Given the selective omission of closed-class items, the first possibility to be checked was that maybe children do not learn at this stage closed-class or 'function' words. To test this possibility, Brown (1973) searched through child corpora available to him, but the hypothesis was wrong: He did find closed-class or function words in children's two-word and early multiword speech, among them *more, no, off* and the pronouns *I, you,* and *it.* In fact, most of what Braine (1963) called pivot-open combinations were built on closed-class items as pivots.

As children do know and use the closed-class words in other types of word-combinations, this leaves two options, highly relevant for our original question for this study. One possibility is that these are productive word-combinations in which children express the single content word carrying the gist of the meaning, as proposed by Ninio (2006). As children are perfectly able to produce word-combinations with closed-class items, it seems a matter of choice (as well as the violation of a marginal rule of grammar) not to include them in their utterances at a stage when their ability to generate a sentence of more than two or three words is as yet shaky. Similar to the wording of telegraphs where you pay by word, when inclusion of words in the message is costly in terms of processing effort, children will not generate a longer sentence including functional words if they are not essential for conveying the gist of the message. The words 'missing' from the utterances may have important grammatical functions in the relevant adult sentences, but the words 'retained' are the substantive words carrying the semantic content of their respective phrases. Thus, 'telegraphic speech' represents an extremely elliptical method for satisfying the semantic and syntactic requirements (valency) of the predicates around which the sentence is built – but satisfying them nevertheless. The word-combinations correctly 'project' the lexical valency of the predicate words involved, satisfying both semantic and syntactic requirements. For example, the 'telegraphic' sentences *"want bottle"* *or "bring chair"* cited above satisfy the semantic requirement of the verbs *want*

and *bring* for an object-type logical argument, the former for the thing wanted, the latter for the thing to be brought; the child-speaker even has the correct idea where to place them relative to the verb, meaning that he already has a workable syntactic valency-frame established for this verb, including the VO word order for the verb and the direct-object elements. There is a rule that this sentence is breaking to do with the obligatory determinants heading noun-phrases in English, but at the bottom line, that rule is irrelevant for satisfying the valency requirements of the verb *want* or *bring,* and that is what 'telegraphic' sentences appear to take as a first priority.

The second possibility is that 'telegraphic speech' is imitated unanalyzed input strings, from which function words are omitted because of their low perceptual salience; being unstressed words, function-words may not actually be heard by the child (Gleitman & Wanner, 1982). This alternative will be tested by looking at some new data, asking whether the same children use function words in other constructions in their speech.

Introduction to the present study

In the following, we shall focus on children's use of bare nouns in their earliest word-combinations, and attempt to answer through them the question whether their entry into multiword speech consists of imitating frozen chunks or else immediately productive combinations of predicates and their arguments. We are focusing on the verb-direct object relation that appears to elicit a tendency to use bare common nouns in two-word speech.

We shall carry on two studies. In Study 1 we shall examine the earliest verb-object combinations generated by a large sample of English-speaking children. Our goal is to determine how general the phenomenon of 'telegraphic speech' is; in particular, we will ask how frequent are bare common nouns in children's early two-word speech. Although the characterization of children's two-word speech as 'telegraphic speech' is well-accepted, the claim that children produce bare nouns as complements of verbs should be checked on large samples in order to make generalization possible. Thus the first study will closely examine a large sample of English-speaking children's earliest and shortest verb-direct object combinations. We will examine the distribution of different nominal complements serving as direct objects in children's speech. The goal is to determine if bare common nouns are used by a majority of children, and what is their proportion in children's speech. Then, we shall compare children's direct objects to parents' objects in equally short verb-object sentences in a large corpus of child-directed speech, asking whether it is likely that children learn their sentences with bare common nouns from parents' input sentences as frozen chunks.

In a second study, we shall ask whether the bare common nouns produced by these young children reflect their perceptual problems with the acquisition of a vocabulary of closed-class unstressed function words that determiners are.

We shall ask if children's use of bare common noun objects is associated with a general absence of determiners in their connected speech.

Method

Participant children

The study uses a sample of 43 children, taken from the speech observations on English-speaking children available in the CHILDES (Child Language Data Exchange System) archive (MacWhinney & Snow, 1985; MacWhinney, 2000). The CHILDES archive stores the transcribed observations collected in various different research projects, each with its individual population and methodology. We have selected projects among the ones available using the criteria that the observations were of normally developing young children with no diagnosed hearing or speech problems, native speakers of English, the children were at the start of multiword observations, and their speech was produced in the context of naturalistic, dyadic parent-child interaction.

Only spontaneous child utterances were included in the study, excluding immediate imitations of previous parental utterances in the last three turns at speech. For each utterance marked in the original transcriptions as one uttered by the child, we hand-checked the context to made certain that the line was indeed child speech and not, for example, an action description or parental sentence erroneously marked as child speech. The size of each participant child's individual contribution was kept to maximum 300 multiword utterances, starting from the first observation in which they produced multiword utterances. This limit was imposed as it appeared to define the stage of two-word speech in all children. The children's two-word utterances were manually annotated for the verb-object (VO) grammatical relation, based on Hudson (1990) and Quirk, Greenbaum, Leech, & Svartvik (1985). Syntactic annotation was done by two graduate students with training in linguistics, and a blind reliability check revealed that the accuracy of the coders was above 98%.

We only included in the present study children who produced at least 10 tokens of the verb-direct-object expressed in two-word sentences as this number is the minimum necessary for the computation of meaningful proportions from their individual results. This process resulted in the selection of 43 children from 16 research projects in the CHILDES archive: the British projects Howe, Manchester, and Wells, and the American projects Bates, Bloom 1970 and 1973, Brown, Cornell, Higginson, New England, Peters-Wilson, Post, Sachs, Suppes, Tardif and Van Houten (MacWhinney, 2000). Other results were published in Ninio (2011).

As we had hoped, including in the study children's earliest 300 multiword utterances from the start of word-combinations tapped into the two-word period. The Mean Length of Utterance (*MLU*) in words was between 1.20 and 2.27 for the sample, with an average value of 1.66 (*STD* 0.24).

Table 1. Distribution of the expressions serving as direct objects in children's two-word long sentences, by part of speech

Part of speech	Number of tokens		Percent of tokens %	
	Mean	STD	Mean	STD
Pronoun	15.21	13.61	61.88	25.60
Common noun	8.14	8.82	31.66	22.40
Proper noun	0.86	2.18	4.19	13.19
Other	0.53	1.01	2.27	4.30
Total	24.74			

The total number of two-word long sentences produced by the sample was 5,323; the mean number per child was 123.79 with a standard deviation (*STD*) of 80.90. Of these, the total number of tokens of two-word long verb-object sentences was 1,064; the mean number per child was 24.74 (*STD* 18.37). This included a mean of 10.21 different verbs per child (*STD* 5.92). The maximal number of different verbs a child used to generate VO sentences in this period was 25. The mean age of the children generating the VO sentences was 1 year, 10 months and 29 days – that is, almost exactly 1;11 – with a standard deviation of 2 months 8 days.

Results and discussion

Table 1 presents the distribution of the expressions serving as direct objects in the children's sentences, by part of speech.

The majority of children's direct objects were pronouns; together, pronouns and common nouns make up almost all of their direct object expressions. Apart from these, direct objects were also expressed by a few proper nouns, and, rarely, other parts of speech, e.g. nonfinite verbs (as in the sentence *"want see"*), or interjections (as in *"Say 'yeah'"*).

Children's bare nouns as the direct object in two-word long sentences expressing the VO syntactic relation

Table 2 presents examples of sentences with a common noun object produced by the sample.

Among the 43 children of the sample, 38 or 88.4% produced two-word long verb-object sentences with bare common nouns as the direct object. The mean percent of tokens with such objects was 31.7%, with a 22.4 standard deviation. If we pool all tokens of direct objects produced by the sample, there are overall 1,064 VO sentences, and 350 or 32.9% of them have a common-noun object. The

Table 2. Sentences with common noun objects produced by children

Child (project)	Sentence	Age
Adam (Brown)	wash hand	2; 3.04
Allison (Bloom)	open box	1; 10.00
Anne (Manchester)	want cupboard	1; 10.21
April (Higginson)	blow nose	1; 10.0
Aran Manchester)	get cake	1; 11.12
Barry (Howe)	push lorry	1; 7.00
Becky (Manchester)	eating tomato	2; 0.14
Benjamin (Wells)	want beer	1; 11.30
Betty (Wells)	wash baby	1; 9.04
c20 (Tardif)	do blocks	1; 9.20
c32 (New England)	hold book	1; 8.11
Carl (Manchester)	turn page	1; 8.22
Dominic (Manchester)	broke train	1; 11.17
Ellen (Wells)	buy birdie	1; 9.00
Elspeth (Wells)	open window	2; 2.29
Eric (Bloom	see machine	1; 9.14
Eve (Brown)	wipe juice	1; 6.16
Gail (Manchester)	fit dolly	1; 11.27
Gary (Wells)	got ball	2; 0.04
Geraldine (Cornell)	take drink	1; 9.00
John (Manchester)	draw horsie	1; 11.15
Jonathan (Wells)	want omelette	1; 11.29
Kalie (Post)	brush pony	1; 8.04
Kent (Bates)	get feet	2; 4.00
Kevin (Howe)	do drawing	2; 0.00
Lew (Post)	get water	2; 2.28
Liz (Manchester)	see ant	2; 0.07
Melanie (Howe)	write list	2; 0.00
Naomi (Sachs)	push recorder	1; 9.26
Neville (Wells)	ring bells	2; 3.00
Nicola (Howe)	post letters	2; 0.00
Nina (Suppes)	roll ball	1; 11.16
Oliver (Howe)	driving car	2; 0.00
Peter (Bloom)	open train	1; 9.08
Salley (Post)	move book	1; 8.16
Sally (Howe)	throw cat	2; 0.00
Sarah (Brown)	take hand	2; 4.10
Warr (Manchester)	bash feet	1; 10.06

pooled and averaged percentages are almost identical, which is helpful when we come to compare the children's results to parents' pooled corpus data.

What is the source of children's bare noun objects?

The question now is, what is the source of children's bare noun objects? All children's bare noun terms referred to a specific entity, hence this use was a grammatical mistake, as the nouns were missing an obligatory determiner. As we mentioned in the Method section, child sentences included in this study were spontaneous, that is, children's sentences with bare nouns could not have been imitations of parental sentences said in the last three turns at speech. Moreover, it is well known that adults on the whole do not make such mistakes when addressing young children, hence it is close to certain that children did not learn and copy such word-combinations verbatim from their parents' past speech. Nevertheless, it is still possible that children observe adults producing bare common noun direct objects, as in some circumstances they are grammatical, and sometimes adults, too, omit the determiners. In order to find out from the empirical data if child-directed two-word speech in models bare common nouns with sufficient frequency to serve as the input to children's bare noun objects, we will turn again to the available large corpus of parental speech. Testing the hypothesis that children learn their two-word long verb-object sentences from parental two-word utterances expressing the same syntactic relation, including sentences with common noun objects, we shall examine parents' two-word long verb-object sentences, comparing their use of bare common nouns to children's in such equally short sentences.

In most previous studies it was simply assumed that parents do not use bare nouns as syntactic complements but usually the studies did not compare children's tendency to generate such expressions to parents' equally short sentences of the same syntactic type. The thing to remember is that it is not impossible for children to learn to produce bare common noun objects from parents' two-word long speech. Bare nouns are not grammatical errors in some cases such as mass nouns, generic plural nouns and so forth; and, in addition, parents addressing young children may occasionally omit function words such as determiners even though they are formally required. In the absence of detailed examination of parents two-word speech, we cannot rule out the possibility that such short input sentences could serve as models for children's bare nouns in a verb-noun combination; that is, it may still be possible that children learn such elliptical expressions from their parents' speech of equal length. We need to find out whether the proportion of bare nouns produced by children is or is not significantly higher than produced by parents in equally short sentences. If the distributions are similar, we may consider two-words long input sentences as the functional input for children's two-word sentences, whether frozen or productive. If the distributions are dissimilar, the two-word

parental input speech may still be the input for productive combining rules, but not for the taking over of frozen chunks from parental two-word to child two-word utterances.

To remind us, we concentrate on common-noun direct objects because it is a construction we do not expect to be generated very often by adult speakers of English. If this is a correct generalization, namely, if we find that parents indeed produce a much lower proportion of bare common noun objects in their two-word long sentences than children do in theirs, this would lower considerably the likelihood that children's two-word speech is the copy of parents' input sentences, taken over as frozen unanalyzed chunks.

Comparing children's two-word long sentences expressing the VO to parents' two-word VO sentences

As we mentioned in the Introduction to this paper, a recent study revealed that English-speaking parents' two-word utterances expressing the three core syntactic relations, namely, the subject-verb, verb-object and verb-indirect object relations, express the verbs' nominal complement in a very large proportion of cases with the help of pronouns and determiners, and sometimes with proper nouns, but very few are expressed by bare common nouns: only 4.1% of sentences with a subject, object or indirect object employed bare nouns (Ninio, 2014). These results are suggestive, but a more accurate comparison would be to compare children's two-word long verb-object sentences specifically with parents' two-word sentences expressing the same verb-object syntactic relation. As a background for this comparison, a short description of the parental corpus is in order.

A corpus of parents' child directed speech

As an estimate of child-directed speech, we are using an already constructed large corpus or collection of transcribed sentences, representing English-language parental child-directed speech which was built and annotated in a previous stage of this study (Ninio, 2011). This corpus represents the linguistic input that young children receive when acquiring syntax. We systematically sampled the English transcripts in the CHILDES archive (MacWhinney, 2000) for parental speech. The use of pooled corpora of unrelated parents as a representation of the linguistic input is a relatively conventional move in child language research (e.g., Goodman, Dale, & Li, 2008). Multiple speakers of child-directed speech may provide a good estimate of the total linguistic input to which children are exposed, which includes, besides the speech of the individual mother or father, also the speech of grandparents, aunts and uncles, older siblings and other family members, neighbours, care professionals, and so forth, represented in our corpus by the speech of mothers and fathers unrelated to the individual child. The pooled database represents the language behaviour exhibited by the community as a whole when addressing young children. The building of the corpora followed

closely the principles established in linguistics for constructing systematically assembled large corpora (Francis & Kučera, 1979).

Included in the corpus were parents addressing normally developing young children, native speakers of English, their speech produced in the context of naturalistic, dyadic parent-child interaction. Each parent was selected individually, so that from the same research project involving the same target child, either the mother, or the father, or both parents were included as separate speakers, as long as either or both passed the criteria for inclusion. This process resulted in the selection of 506 parents from 33 research projects in the CHILDES archive: the British projects Belfast, Howe, Korman, Manchester, and Wells, and the American projects Bates, Bernstein-Ratner, Bliss, Bloom 1970 and 1973, Brent, Brown, Clark, Cornell, Demetras, Feldman, Gleason, Harvard Home-School, Higginson, Kuczaj, MacWhinney, McMillan, Morisset, New England, Peters-Wilson, Post, Rollins, Sachs, Suppes, Tardif, Valian, Van Houten, and Warren-Leubecker (MacWhinney, 2000).

In order to avoid severely unequal contributions to the pooled corpus, the number of utterances included from each parent was restricted to a maximum of 3,000. All transcribed dialogue and the action and other contextual comments were checked in order to ascertain that we include only spontaneous utterances from target parent to target child. The mean age of the children addressed was 2.25 years. The sentences were manually annotated for core syntactic relations. Blind recoding showed that coding reliability was close to 100%. For more details of the corpus-building and the coding process performed at the previous stage of the study, see Chapter 2 of Ninio (2011).

Results of the comparison of parental two-word verb-object sentences with children's

In the pooled corpus of child-directed speech, there were 4,180 tokens of two-word long utterances expressing the VO relation. The great majority of the direct objects (2,909) were pronouns, making up 69.6% of the total. Only in 261 of the sentences, namely, 6.2% of the total, were the direct objects expressed by bare common nouns. This is a much smaller percentage of sentences than children's 32.9% of sentences with common nouns we found above.

In addition, we computed the proportion of parents who expressed direct objects by common nouns. Out of 417 parents with two-word long VO sentences, 116 or 27.8% produced at least one common-noun object; by comparison, 88.4% of the child sample produced such objects.

It appears children's distribution of complements does not copy parents', in particular in the extensive use of bare nouns as the single-word complements of verbs. Even in absolute terms, adults use in two-word child-directed speech a very small amount of bare common nouns, making it unlikely that children's usage is an adoption of adult sentences as frozen chunks. These data are more supporting

of the hypothesis that children learn from parental two-word sentences productive rules for expressing a verb and its semantic-logical object.

Study 2. Children's use of closed-class determiners

Although we found that parents' two-word sentences contain very few bare common noun objects, this does not by itself provides decisive evidence regarding the productivity or frozenness of children's two-word speech. Besides the possibility that these are productive word-combinations, there is a possibility that verbs with bare noun objects are frozen chunks which are learned from longer parental sentences by omitting the determiners. According to the perceptual-deficit conceptualization of 'telegraphic speech' (Gleitman & Wanner, 1982), children may perceptually miss the unstressed determiners in longer utterances, thus it could be that they learn *"Open box"* as a **frozen** expression from such a parental sentence as *"Open the box"*. In order to test this possibility, we shall examine the children's early word combinations for two-word sentences consisting of a closed-class determiner (articles *a(n), the, some*; demonstratives *this, that*; possessive pronouns *my your*, and so on) and a common noun, comparing their frequency to the verb-bare noun combinations produced by the same children at the same period. This is similar to Brown's (1973) strategy in the study mentioned above, in which he searched for evidence that children do not learn functional words at this developmental stage, with the difference that we are specifically concentrating on the shortest multiword utterances, namely, two-word long ones. We are expecting that to the extent that children generate bare-noun objects as frozen because of a perceptual problem to hear the unstressed determiners, they will not produce, at the same stage of development, word combinations consisting of the same unstressed determiners and common nouns.

Method

We have examined the speech of the same sample of 43 children as before, most of whom produced bare common noun direct objects. We included determiner phrases only with the closed-class determiners *a, all, an, another, any, five, her, his, many, mine, more, my, no, not, one, our, some, that, the, this, those, three, two* and *your*, excluding possessive determiners using proper names such as *"Anne's"*, with or without the possessive *'s*.

Results and Discussion

Table 3 presents the number of two-word sentences produced by the children of the sample with, respectively, closed-class determiners and common nouns, and verbs with bare common noun direct objects.

Table 3. Number of two-word sentences produced by children containing closed-class determiners and common nouns, and verbs with bare common noun direct objects Sentences with common noun objects produced by children

Child (project)	Determiner Phrase with closed-class determiner	Verb-Object combination with bare noun object
Adam (Brown)	0	37
Allison (Bloom)	12	13
Anne (Manchester)	60	14
April (Higginson)	26	3
Aran Manchester)	23	11
Barry (Howe)	24	3
Becky (Manchester)	98	20
Benjamin (Wells)	15	7
Betty (Wells)	5	3
c08 (Van Houten)	5	0
c20 (Tardif)	8	2
c32 (New England)	6	1
Carl (Manchester)	41	5
Dominic (Manchester)	57	6
Ellen (Wells)	12	5
Elspeth (Wells)	7	4
Eric (Bloom	24	7
Eve (Brown)	0	11
Faye (Howe)	19	0
Gail (Manchester)	3	3
Gary (Wells)	12	3
Geraldine (Cornell)	86	15
John (Manchester)	71	32
Jonathan (Wells)	12	7
June (Higgins)	124	0
Kalie (Post)	21	3
Kent (Bates)	4	5
Kevin (Howe)	11	29
Lew (Post)	10	5
Liz (Manchester)	49	18
Melanie (Howe)	3	6
Naomi (Sachs)	27	10

Neville (Wells)	6	8
Nicola (Howe)	15	2
Nina (Suppes)	124	2
Oliver (Howe)	33	5
Peter (Bloom)	13	10
Philip (Howe)	11	0
Salley (Post)	11	2
Sally (Howe)	22	2
Sarah (Brown)	0	24
Seth (Peters-Wilson)	24	0
Warr (Manchester)	184	7
Mean	**30.7**	**8.2**
Standard Deviation	39.4	8.8

The comparison reveals that children produce many more determiner phrases with closed-class determiners than they produce bare common noun verb-object combinations. A two-tailed, paired *t*-test revealed that the difference is highly significant: $t(42) = 3.71$, $p < 0.001$. That is, children are much more likely to generate determiner phrases using closed-class determiners than they are to drop the same unstressed determiners from parental verb-object phrases which they are thought to copy in a frozen manner.

In addition, we tested the probability that children with a larger number of bare noun objects generate a lower number of determiner phrases with closed-class determiners. A Pearson correlation coefficient was $r(41) = 0.0592$, $p > 0.05$. The results demonstrate that there is no negative correlation between the use of closed-class demonstratives and the use of bare noun complements, rejecting the hypothesis that the more children ignore determiners in copying parental verb-object sentences, the less they generate determiner-noun phrases. There is in fact no relation between the production of these two constructions.

As these children were the same age when they generated the determiner-noun phrases as they were when they produced the VO sentences analyzed above, this means that even if they got the idea for bare common noun objects from longer input sentences, this was not the result of a perceptual problem hearing the closed-class determiners. Children can still learn the rule for two-word VO sentences from longer input sentences, but the rule will be the same as the one learnable from two-word long input sentences: Express the verb, plus express its semantic object argument by a single word best expressing the object. Often this will be a pronoun or demonstrative, expressing the argument by a **variable** expression getting its value from the context; but some of the time it

will be a content-word, expressing some **specific** value of the argument. In all probability, these rules map the predicate-argument relations of verbs and their semantic arguments to head-dependent syntax, with the operation Merge or Dependency serving for syntactic connectivity (see also McCune, 2013; Ninio, 2006). The rules children use are productive rules; they may be influenced by children's performance limitations, but not reflecting a misunderstanding, frozen copying or the operation of any other non-syntactic principles in children's early syntactic system.

Acknowledgments

References

Arnon, I. (2010). *Starting Big: The Role of Multi-Word Phrases in Language Learning and Use.* Unpublished Dissertation, Stanford University.

Arnon, I. (2011). Relative clause acquisition in Hebrew and the learning of constructions. In E. Kidd (Ed.), *The Acquisition of Relative Clauses: Processing, Typology, and Function* (pp. 81-106). Amsterdam: John Benjamins.

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

Bloom, L. (1970). *Language Development: Form and Function in Emerging Grammars.* Cambridge, MA: MIT Press.

Bloom, L. & Lahey, M. (1978). *Language Development and Language Disorders.* New York: John Wiley.

Bowerman, M. (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* (pp. 99-179). Baltimore, MD: University Park Press.

Bowerman, M. (1978). Words and sentences: Uniformity, individual variation, and shifts over time in patterns of acquisition. In F.D. Minifie & L.L. Lloyd (Eds.), *Communicative and Cognitive Abilities – Early Behavioral Assessment* (pp. 355-371). Baltimore, MD: University Park Press.

Braine, M.D.S. (1963). The ontogeny of English phrase structure: The first phase. *Language, 39,* 1-13.

Braine, M.D.S. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development, 41*, 164.

Brown, R. (1973). A *First Language: The Early Stages*. Cambridge, MA: Harvard University Press.

Brown, R. & Bellugi, U. (1964). Three processes in the child's acquisition of syntax. *Harvard Educational Review, 34*, 133-151.

Brown, R. & Fraser, C. (1963). The acquisition of syntax. In C.N. Cofer & B.S. Musgrave (Eds.), *Verbal Behavior and Learning: Problems and Processes* (pp. 158-197). New York: McGraw-Hill.

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

Deuchar, M. & Quay, S. (2000). *Bilingual Acquisition: Theoretical Implications of a Case Study*. Oxford: Oxford University Press.

Diessel, H. (2013). Construction Grammar and first language acquisition. In G. Trousdale & T. Hoffmann (Eds.), *The Oxford handbook of Construction Grammar* (pp. 347-364). Oxford: Oxford University Press.

Ellis, N.C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. Doughty & M.H. Long, (Eds.), *Handbook of Second Language Acquisition* (pp. 33-68). Oxford: Blackwell.

Ervin-Tripp, S.M. (1966). Language development. In L.W. Hoffman & M.L. Hoffman (Eds.), *Review of Child Development Research*. Vol. 2 (pp. 55-105). New York: Russell Sage Foundation.

Francis, W.N. & Kučera, H. (1979). *Brown Corpus Manual of Information to Accompany a Standard Corpus of Present – Day American English, Revised and Amplified*. Providence, RI: Brown University, Department of Linguistics.

Gentner, D. (1978). On relational meaning: the acquisition of verb meaning. *Child Development, 49*, 988-998.

Gleitman, L. & Wanner, E. (1982). Language acquisition: the state of the state of the art. In E. Wanner & L. Gleitman (Eds.), *Language Acquisition: The State of the Art* (pp. 1-50). Cambridge, MA: Cambridge University Press.

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

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

Lieven, E.V.M., Pine, J.M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language, 24,* 187-219.

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language, 30*, 333-370.

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

Macnamara, J. (1972). Cognitive basis of language learning in infants. *Psychological Review, 79*, 1-14.

MacWhinney, B. (1982). Basic syntactic processes. In S.A. Kuczaj (Ed.), *Language Development:* Vol. 1. *Syntax and Semantics* (pp. 73-136). Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd ed.). Mahwah, NJ: Erlbaum.

MacWhinney, B. & Snow, C. (1985). The child language data exchange system. *Journal of Child Language*, *12*, 271-295.

McCune, L. (2013). *Developing a Linguistic Medium: From Words for Things to Words for Grammar.* Unpublished manuscript, Rutgers University, NJ.

McCune, L. & Vihman, M.M. (1999). Relational words and motion events: A universal bootstrap to syntax. Paper presented at the SRCD Biennial Meeting, Albuquerque, NM.

McCune-Nicolich, L. (1981). The cognitive bases of early relational words. *Journal of Child Language, 8,* 15-36.

Ninio, A. (1988). On formal grammatical categories in early child language. In Y. Levy, I.M. Schlesinger, & M.D.S. Braine (Eds.), *Categories and Processes in Language Acquisition* (pp. 99-119). Hillsdale, NJ: Lawrence Erlbaum.

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

Ninio, A. (2006). *Language and the Learning Curve: A New Theory of Syntactic Development.* Oxford: O.U.P.

Ninio, A. (2011). *Syntactic Development, Its Input and Output.* Oxford: O.U.P.

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

Peirce, C.S. (1865/1982). Logic of the sciences. In M. Fisch, C. Kloesel, E. Moore, N. Houser et al. (Eds.), *The Writings of Charles S. Peirce: A Chronological Edition.* Vol. 1 (pp. 322-336). Bloomington, IN: Indiana University Press.

Peters, A.M. (1983). *The Units of Language Acquisition.* Cambridge: Cambridge University Press.

Pine, J. M. & Lieven, E.V.M. (1993). Reanalyzing rote-learned phrases: Individual differences in the transition to multi-word speech. *Journal of Child Language, 20*, 551-571.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language.* London: Longman.

Reboul, A. (2001). Foundations of reference and predication. In M. Haspelmath, E. König, W. Oesterreicher, & W. Raible (Eds.), *Language Typology and Language Universals: An International Handbook.* Vol. 1 (pp. 509-522). Berlin: De Gruyter.

Tesnière, L. (1959). *Elements de Syntaxe Structurale.* Paris: Klincksieck.

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