**Learning from examples in syntactic development**

Anat Ninio

The Hebrew University of Jerusalem \*

Keynote address at the 11th Australasian Human Development Conference,

Sydney University, Sydney, Australia, July 9, 1999.

\*Portions of the research reported here were also presented at two meetings of the Child Language Research Group, Sydney, Australia, 19 February, 1999 and 30 April, 1999; at the Colloquium of the Social-Developmental Group, Macquarie University, Sydney, Australia, 2 March, 1999; the Psychology Department Colloquia, Macquarie University, Sydney, Australia, 9 August, 1999, and at an invited talk, Keio University and Ochanomizu Women's University, Tokyo, Japan, September 10, 1999. Thanks are due to Mike Tomasello for his helpfulness and patience in elucidating various complications of the Travis corpus; to Moti Rimor for permission to use his observations of Ruti; to Naomi Bacon, Lucy Bahat, Eli Berkowitz, Adi Brill, Sarit Dag, Rachel Hadani, Noa Hocherman, Margalit Kazins, Tamar Keren-Portnoy, Chagit Magen, Liat Malhi, Sharon Marbach, Noga Meir, Ori Nahari, Ofri Tal, Ariel Wasserteil, Gary Weinstein, Chamutal Yakir and Osnat Yogev who collected the longitudinal language corpora.

The research reported in this paper was supported by a grant from The Spencer Foundation. The data presented, the statements made, and the views expressed are solely the responsibility of the author.

Address for correspondence: Department of Psychology, The Hebrew University, Jerusalem 91905, Israel. E-mail address: msninio@mscc.huji.ac.il

Learning from examples in syntactic development

ABSTRACT

It is usually assumed that children have to learn abstractions such as form-class categories (noun, verb), or grammatical relations (subject, direct object). It seems an impossible task. I want to suggest that children do not learn abstractions at all but concrete examples. They generalize these to other items by analogy. This system works efficiently because the first items learned are usually the 'best models' for the whole domain. Linguistic research shows that in natural language, frequency may be inherently correlated with generality, prototypicality or simplicity. As there is a well-known effect of input frequency on the order of learning, it brings children to learn the best exemplars of a domain first. Developmental research has accumulated many examples of children learning the most prototypical as well as most frequent items first, in various domains such as the lexicon, syntactic rules or morphology. The concrete knowledge acquired about the first-learned items can be transferred to other less central items in the same fields because the first-learned are general, and the other items are their special cases. The inherent structure of linguistic categories seems to insure that language is easily learnable from the linguistic input.

I want to thank the organizers of this conference for inviting me to give this talk. I am glad of the opportunity to present a concept of language development and maybe cognitive development in general that deserves very serious consideration.

The central issue to explain for any model of language development is adults' undoubted ability to generalize old rules to new items. Let's assume that the year is around 1970, and we just heard for the first time a sentence using a new verb 'to fax'. It went on something like 'I'll fax you that page tomorrow'. From that moment on we were able to use the new verb 'fax' in all the sentences we would normally use some word like 'send' or 'mail', for instance, "She forgot she was supposed to fax the results to our group". It is not necessary to learn each rule afresh for this verb, or, more strongly, it simply isn't true that people need to learn each rule separately for every verb that they know.

There are roughly speaking two schools of thought regarding the type of knowledge underlying our ability to generalize old rules to new exemplars. One school assumes that we possess abstract rules, defined for abstract symbols. To remain with our 'fax' example, this is a ditransitive verb, getting both a direct object and an indirect object. We are supposed to possess some abstract category symbol for this subcategory of verbs,and a set of abstract rules using this symbol. Once you learn that the newly encountered item like 'fax' belongs to this category, automatically all the rules of the category apply to it. There are abstract rules that regulate what to do with the subject of the sentence, with the verb, with the direct object, and with the indirect object. All this is phrased in completely general terms without mentioning any particular verb or ditransitive verb, old or new.

The question of course is where would such a set of abstractions come from? One possibility is that they are extracted out of the concrete examples a child hears during the process of language acquisition. Let's say that a child would learn how to use a few of these verbs like 'give' and 'send' and 'show', and she will then by a process of abstraction form a concept of 'ditransitive verb', so that the rules of usage are from that moment on phrased not in terms of any concrete item but in terms of the category symbol.

There is a powerful nativist hypothesis that says, we can't learn these abstractions from the linguistic input around us; it is too difficult. We as children hear random snatches of talk around us, and this noisy set of chance examples is a very bad basis for the process of extraction of the correct abstract symbols and rules. Therefore, it goes on, the basic rules, the basic symbols, the basic verb-types -- all this is not learned at all but genetically inherited. This includes word-class categories like noun and verb, and grammatical relations such as subject or direct object, and, basically, all the fundamentals of grammar.

The second school of thought has a rather simpler proposal. It says, we do not operate with abstractions at all. That is not what underlies our ability to generalize our knowledge to new items. Instead, we operate with concrete examples, and learn our rules for these concrete items. When we encounter a new item, we compare it to all previously mastered examples and treat it as the example it most closely resembles. We do not extract abstractions -- we operate by similarity-matching and by analogy.

In my talk today I would like to offer some evidence that bears on this fundamental debate. My data strongly supports the learning hypothesis and implies that there exist basic conditions for an efficient system built on central examples and analogy. My basic point is that children are not exposed to a random collection of examples from which they are supposed to extract abstract generalizations. Language is not like that. It has a structure that assures its own learnability.

I will be arguing for the idea that a central learning process in language development is the learning of 'good examples'. Once a few of these 'good examples' are learned, it is possible that the learning process is over: the learner can then use these 'good models' in order to apply the same rules to less central items learned later.

This can work of course under one condition: if there are things like 'good examples' around, and if children have a reason to learn them first!

I want to claim that both conditions are met. First, different linguistic domains are organized around some exceptional items which are their 'best exemplars'. These items are simpler, more general, and more widely usable than other items. Second, they are also -- and this is crucial -- much more frequent than other items in the same domain.

The existence of these 'good model' and their high frequency is the basic given that insures that children will start learning new rules for the 'best exemplars' of each domain, and these can serve as 'good models' for the other less general items.

Let's start with an example not from the domain of verbs and their subtypes but from another kind of words: size adjectives. There are about ten pairs of size-adjectives in English, like wide-narrow, tall-short, fat-slim and so on. The question is, how do children come to understand the way you are supposed to use these adjectives, their meaning, their funtioning in relation to other words?

We as adults possess a set of general rules that applies to all size-adjectives. Most basically, if we want to combine one of these size-adjectives like 'large' with another word X in order to generate the meaning 'large X', the rule is : put 'large' or another of these size-adjectives first, the the word you want to modify like 'table', together 'large table'. As an abstraction, we get the rule:

Size-adjective X,

namely, put the size-adjective first and the object-name second, and you get the correct meaning-combination.

**So how do children learn this abstract or general rule?**

In the best possible words, they learn it first on a lexical word-specific basis for one or two of this set of size-adjectives, and the items they learn it on are very very general and can serve as 'good model' for the other, more specific words.

This is exactly what happens in acquisition. First, the semantic domain of size-adjectives indeed possesses two central items, which are 'big' and 'little'. These adjectives are very general in meaning; you can use 'big' for 'tall', wide', fat' and so on if you really wanted to. The more specific large-size adjectives are special-cases of these general terms. In formal terms, they are almost-hyponyms of 'big' and 'small'. This makes 'big' into a very good model for all the large-size adjectives, and 'small', for all the small ones. You learn to generate ruleful combinations with 'big' like 'big doll', you can very easily transfer that knowledge to the more specific synonyms and generate by analogy, 'tall doll' and 'wide doll' and so on.

The question is, why would children learn 'big' and 'small' first? There are two interconnected reasons for that. First, it is predicted that they'll learn these words first because they are semantically the simplest of the lot: they relate only to size, but not to the dimension on which the objects are judged like their height, weight and so on. Second, it is predicted that they'll learn them first because they are used very frequently in the speech of the adult caretakers, and the simple frequency-effect assures that they are learned first before less frequent items.

These predictions are correct: 'big' and 'little' are indeed very frequently used by adults speaking to children, and they are really the first size-adjectives children learn, especially to combine with nouns. This is very old and well-established finding in the field, for instance Blewitt (1982).

I want to continue and to claim that this is not an isolated phenomenon but it is equally true for other domains of language, and in particular for various kinds of verbs.

First I want to present some empirical data on what do we mean when we say that 'good models' are frequent in the linguistic input. It is quite dramatic. It is not the mild effect that what we regularly mean by a frequency effect, but rather, there are extreme frequency differences between a few very general verbs and the rest.

Table 1 presents my own frequency data on the speech of 48 Hebrew-speaking mothers addressed to their children who were between 10 months and 32 months of age -- most in the second year of life which is the crucial period for the beginning of multiword combinations in children. This represents more than 80 hours of video-recording of dyadic interaction in the home. The frequency is of different verbs in multiword combinations only, produced by these mothers. I excluded single-word utterances, the question was about multiword sentences that demonstrate to the children how to use these verbs in multiword or syntactic combinations. The first statistics is of transitive verbs of any kind, namely verbs that can get a direct-object.

**Table 1. Transitive verbs in maternal speech (multiword sentences)**

There were a total of 22,931 sentences with some transitive verb.

There were a total of 270 different transitive verbs in these sentences.

Question: how many verbs accounted for half of all sentences?

Answer: 6 verbs accounted for 50% of all sentences.

These were: want, make or do, put, bring, see, give.

The other 263 verbs accounted for the other 50% of all sentences.

There were almost 23 thousand sentences in the observations, so the generalization is

very strong. As we can see, the most frequently used verbs are extremely frequent: the

most frequent 6 verbs and all the other 263 verbs generated an equal number of

sentences, about 11000 sentences!

Similarly for intransitive verbs:

**Table 2. Intransitive verbs in maternal speech (multiword sentences)**

There were a total of 10,266 sentences with some intransitive verb.

There were a total of 189 different transitive verbs in these sentences.

Question: how many verbs accounted for half of all sentences?

Answer: 2 verbs accounted for more than 50% of all sentences.

These were: come and go

The other 187 verbs accounted for the other 50% of all sentences.

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

The picture is even more extreme with intransitive verbs: just two verbs account for half of all multiword sentences involving such verbs, which is about 5,000 sentences for these two verbs, and the same amount for the other 187 verbs.

I don't want to show the whole table of distribution but of course very many verbs generated just one or two sentences each, in the combined sample of 48 mothers. There were other verbs also that never occurred in 80-plus hours of observation -- all together the maternal input covered a very small number of different verbs in the Hebrew language! If we think about the many thousands of verbs which never surfaced in the input, and the fact that the very-frequent ones occurred every couple of minutes -- we can start to get a feel for the frequency effect that operates in this speech sample.

The verbs that occur with the very high frequency in maternal speech are very general, semantically very simple and wide-applicable; and in general can be seen as kinds of pro- forms for the other verbs, just like 'big' and 'little' relate to the other size-adjectives.

Next, let us examine the acquisition data. Here the sample is 16 children who were followed longitudinally and we documented the first verbs they used in a syntactic verb-object combination. These children are unrelated to the mothers' sample, so there is no possibility of any direct effect of the children on the mothers' speech. One of the children was actually acquiring English!

**Table 3. Distribution of the first two verbs appearing in VO combinations in the sample**

**(N=16)**

|  |  |  |
| --- | --- | --- |
| **Verbs** | | **Number of children** |
| **Hebrew** | **English** | **(N=16)** |
| raca | want | 13 |
| lakax | take | 4 |
| natan | give | 4 |
| asa | make/do | 3 |
| hevi | bring | 2 |
| -- | find | 1 |
| -- | get | 1 |
| ra'a | see | 1 |
| shama | hear | 1 |
| axal | eat | 1 |
| shata | drink | 1 |

The next table presents for the same children the first two verbs occurring in subject-

verb-object syntactic combinations.

**Table 4. Distribution of the first two verbs appearing in SVO combinations in the sample**

**(N=13)**

|  |  |  |
| --- | --- | --- |
| **Verbs** | | **Number of children** |
| **Hebrew** | **English** | **(N=13)** |
| raca | want | 9 |
| asa | make/do | 6 |
| axal | eat | 2 |
| shata | drink | 1 |
| hexin | prepare | 1 |
| bana | build | 1 |
| ciyer | draw | 1 |
| lakax | take | 1 |
| sam | put | 1 |
| sagar | close | 1 |
| -- | ride | 1 |
| yexol | can | 1 |
| marshe | allow | 1 |

The next table presents the first two intransitive verbs in any kind of word-combination,

for a sample of 20 children, including the ones we saw before and a few more.

**Table 5. Distribution of the first intransitive verbs appearing in word-combinations in children (N=20)**

|  |  |  |  |
| --- | --- | --- | --- |
| Verbs | | # children (N=20) | |
| Hebrew (stem) | English gloss | 1st | First 2 |
| ba | come | 13 | 14 |
| nafal | fall | 2 | 7 |
| halak | go | 2 | 5 |
| yashav | sit | 1 | 2 |
| yashan | sleep | 1 | 1 |
| `af | fly | 1 | 2 |
| zaz | move | 0 | 4 |
| baka | cry | 0 | 1 |
| kaav | hurt | 0 | 1 |
| kam | get up | 0 | 1 |
| yaca | exit | 0 | 1 |
| tas | fly (plane) | 0 | 1 |

The verbs first acquired by children are with a very high probability from among the high-

frequency, highly general verbs modelled by mothers when they talk to young children. Among

the transitive verbs these are 'want' and 'make', 'give' and 'take' and the like; and among the

intransitive verbs, it is 'come' and 'go'. More specific verbs have a greatly reduced probability to

be among the first two verbs acquired in combination by children.

The last question is, is there are evidence for the earliest verbs facilitating the acquisition of the

later, more specific verbs in syntactic combinations?

The answer is positive. We plotted the development of verb-object and subject-verb-object word-

combinations in two children, Ruti and Travis, as a function of age. The dependent measure is

the cumulative number of different verbs participating in each type of construction. Each verb is

counted at the age when it is first produced in the relevant syntactic construction.

In both children, both cumulative series show a rising exponential or geometrical function,

starting very slowly and accelerating gradually. The two graphs of a given child are almost

identical in their shape. It is evident that these graphs have the characteristic shape of typical

gradually accelerating learning curves: The time it takes to apply the new rule to yet another

verb is much longer at the beginning of acquisition of that rule, and it gets shorter the more verbs

the children have already learned to produce in the relevant pattern. The fact that the same

speed-up of acquisition occurs both in VO and SVO is one of the central pieces of evidence

suggesting that the speed-up is specifically tied to the number of previously produced verbs in the

same kind of combination. There is apparently a great deal of facilitation or generalisation from

one verb to another in the process of learning a new combinatorial rule.

Inspection of the learning curves in the two figures reveals that most of the massive facilitation is

provided by the first and second verbs in each type of word-combination, and that even the third

verb already has a very small additional facilitating effect. This implies that most of the general

or abstract knowledge about the VO and the SVO positional patterns are acquired in the context

of the first two combining verbs in each pattern.

Apparently, breaking into a new syntactic combination means solving the conceptual problems

associated with that pattern once and for all. This puts a heavy burden on the first two verbs, as

they must provide strongly prototypical instances of the relevant combinatorial patterns. As we

saw, the first two verbs are indeed the most general verbs possible in the relevant construction.

CONCLUSION

The results of this study suggest that children learn new combinatorial rules first for a few verbs

in a piecemeal way, but immediately begin transferring some more general and abstract principle

to other verbs so that applying the same combinatory principle to new verbs becomes

progressively easier. This process is equivalent to the gradual consolidation of an abstract

grammatical relation such as the verb-object relation, as well as to the consolidation of a

similarity-class of verbs to which the relevant principle applies, namely, a lexical form-class

which is relative to, and specific to, the syntactic rule applying to its members.

Apparently, the source of the generalisable knowledge is the first two or three verbs that combine

in a novel syntactic pattern. These verbs tend to be generic verbs that express the relevant

combinatorial property in a relatively undiluted fashion. Thus the earliest lexical-specific

transitive concepts are the most general lexical concepts possible. The specific pathbreaking

verbs may vary with each major step in syntactic development; for each step there may be some

verbs which represent the most appropriate prototype for the relevant syntactic combination.

These verbs break the path for other verbs to follow without having to undergo the same difficult

process of learning everything from scratch.

In all, the results of these studies on the beginning of syntactic combinations in children provides

quite strong support for a model of learning in which children learn the best exemplars of a

domain first. The concrete knowledge acquired about the first-learned items can be transferred to

other less central items in the same fields because the first-learned are general, and the other

items are their special cases.

The learning process for linguistic rules of all kinds suggested by these results is extremely

similar to concept formation: apparently while learning the features of a few concrete instances,

children form what is the functional equivalent of abstract concepts applying to kinds of objects

of the same type. This concept-formation process is facilitated, or even made possible by the

statistical structure of language, by which the simplest, most general, and most model-like

instances in each domain are the most frequently used instances of that domain. The well-

documented input-frequency effect thus has a constitutive role in language acquisition which is

cardinally different from its supposedly "behavioristic" character: It makes possible the formation

of abstract concepts, on the basis of prototypical or "best model" instances of a linguistic domain.

At the same time, the proposed acquisition mechanism is a proper learning procedure, without the

need to invoke genetically inherited linguistic concepts or any nativist notions of this kind.

Human languages are robust systems with functionally shaped features; such a system would have

precisely the characteristics that makes it the easiest possible to learn from the available linguistic

input. The correlation of use frequency and simplicity, generality and best-model characteristics demonstrated in this study, is one of the optimal design-characteristics of language that make its

learning by children a highly solvable cognitive problem.

**The intellectual context of this proposal**

The proposed learning mechanism fits in very well into the current school of thought in cognitive psychology as well as in artificial intelligence that search for concrete, example-based alternatives to abstract thought and learning. The beginning of this approach is without doubt in Rosch's seminal work on prototype-centered categories in human cognition. Recent resurgence of interest in, and developments of, Prototype Theory as well as some adjoining fields such as problem-solving demonstrate that the time is ripe to re-think language development in non-abstract terms.

The major intellectual context of the proposed learning mechanism is the following:

A. Prototype Theory (eg., Rosch, 1973, 1975, 1977; Lakoff, 1982, 1987; Taylor, 1989, 1998)

B. Problem-Solving by Analogical Reasoning (eg., Gick & Holyoak, 1983; McAndrews &

Moskowitch, 1985; Perruchet & Pacteau, 1990)

C. Similarity Matching vs. Abstract Rules in Categorization (eg., Gentner & Medina, 1998; Hahn

& Chater, 1998; Nosofsky, 1984, 1992)

D. Experience-Based Computer Learning in Artificial Intelligence (eg., Aka, 1997; Bod, 1999;

Mitchel, 1990).

It seems that the present Zeitgeist or spirit of the times is more than sympathetic to a shifting of

emphasis from abstractions and rules and features to the exemplar, the concrete, and the holistic

in the cognitive field. It is obvious that much empirical research is needed before we can be

confident that the new concrete approaches such as the proposed "Model Learning" are strong

alternatives to the currently popular abstraction-based models of human thought and learning.

However the signs are promising and this kind of work may well gain much support in the

coming decade.