ON THE GRAMMAR AND MODEL-THEORETIC SEMANTICS OF CHILDREN'S NOUN PHRASES

by

Patrick Suppes

TECHNICAL REPORT NO. 181

November 29, 1971

PSYCHOLOGY AND EDUCATION SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

On the Grammar and Model-Theoretic Semantics of Children's Noun Phrases[1]

Patrick Suppes

Institute for Mathematical Studies in the Social Sciences
Stanford University
Stanford, California 94305

I had originally intended to prepare for this colloquium a detailed analysis of noun phrases used by children in English, French and Chinese. That program of work is well under way, but is not sufficiently complete to offer a systematic and summary presentation of results at this time. However, in view of the difficulties I have encountered on other occasions in communicating the ideas of model-theoretic semantics to linguists or psychologists not primarily interested in or familiar with the work in the theory of models in modern logic, I think a discursive and informally organized explanatory paper may actually serve a useful purpose.

At the Institute for Mathematical Studies in the Social Sciences at Stanford we have under way a detailed analysis of several large corpora. Two of them are extensive recordings of children whose first language is English and whose ages are between two and three years. We have 20 hours recorded for one girl and more than 40 hours for a second, with the second still continuing. A third corpus is that of a young French boy, whose age is in the same range as that of the two American girls. With more than 16 hours recorded and transcribed, the data collection for the French boy continues at the rate of one hour per week. Finally, we have recordings from two Mandarin-speaking children, who are also between two and three years of age.

The corpora of all the children are recorded on tape and then transcribed and input into our computer system at the Institute for extensive analysis by a variety of programs. This work is being conducted in conjunction with a number of younger colleagues, and detailed results of the work will be presented in collaborative publications with them. The work in French is being conducted in collaboration with Madeleine Léveillé of the Laboratoire de Psychologie in Paris, the Chinese corpus is being collected and analyzed in collaboration with Dr. Teresa Cheng of the Phonology Laboratory, University of California at Berkeley, and the analysis of the two English corpora, together with all of the computer programming, is being done in collaboration with Mr. Robert Smith of the Institute.

Our objective is to provide a relatively complete grammar and model-theoretic semantics of these corpora. In previous papers (Suppes, 1970 and 1971) I have elaborated on the technical details of the work. The first of these papers describes the methods we are using for constructing probabilistic grammars and the second the model-theoretic approach to semantics. Further application of the notion of probabilistic grammar was made in Elizabeth Gammon's dissertation (1970).

I shall not try to recapitulate the technical details but rather try to explain in an informal way the underlying ideas and their sources.

In the case of the grammar, the analysis is done within a generative framework. The line of attack is to write a generative grammar and to attach to each production rule of the grammar a conditional probability of its use, given that a rewrite occurs of the nonterminal symbol that is the first half of the rule. The applications thus far have been in terms of context-free grammars, but the basic idea is not restricted to context-free grammars. It is certainly applicable in direct fashion to indexed grammars that are context-sensitive but not context-free, and also to optional transformations. Once such grammars are constructed for a corpus and the probabilities for the use of a given rule are estimated by standard statistical methods, an ordinary criterion of goodness-of-fit test can be performed in order to compare one grammar with another for the same corpus. The idea that is new is the introduction of probabilities and the application of standard goodness-of-fit tests to evaluate the grammar. There is more to be said here than this sketch conveys, and I know from previous discussions that a detailed clarification of what is involved in constructing the probabilistic part of such a grammar would be desirable.

However, making a generative grammar probabilistic is a minor affair compared to the difficulties and subtleties involved in adding a model-theoretic semantics to that generative grammar. I therefore want to spend most of my time today discussing in the framework of generative grammars the approach to semantics that grows out of the main thrust of work in mathematical logic.

The technical apparatus of contemporary model theory in logic is substantial, but the underlying ideas, which go back to Frege in the 19th century, are completely intuitive and straightforward. The idea is to have a clear and definite procedure for assigning a meaning to an utterance, and to do this, one must be able to show how each word in a sentence performs a definite function. I admit at once that this statement sounds far too vague and uninformative, but the word _function_ means something more here. As a first recast of this idea, we can begin by saying that we shall use standard techniques of modern mathematics to give a set-theoretical account of the meaning of a sentence. This means that we talk about objects as set-theoretical objects, and thus, we talk about individuals, classes of individuals, classes of classes of individuals, relations between individuals, relations between classes and individuals, etc., functions of individuals, functions of classes, etc. All of these objects are built up in a natural way into a hierarchy of sets, with of course in the classical view, relations and functions being particular kinds of sets, so that when we talk about the meaning of a sentence we must assign to each word a set-theoretical object. In the case of a noun like _men_ we assign the class of men; in the case of an adjective like _green_ we assign the class of green objects. Thus, fairly simple ideas of reference work. Already, however, there are adjectives that create problems. If we think of the phrase _alleged dictators_, it would not do to assign to the adjective _alleged_ the class of _alleged things_, or at least, this already seems to be somewhat strange. Once we leave adjectives and nouns, the picture can become complicated rather quickly. For example, ordinary and simple-minded ideas of reference do not give us any clues of what object to assign to the definite article _the_, or what object to assign to a preposition like _of_. It is for situations like this that set theory was created. The definite article or a preposition do not designate a simple set of individuals, but are more complicated set-theoretical functions or relations. We shall look at some examples shortly.

Another point that needs to be clarified early is that in first approximation it is often easier to assign a meaning that is a set-theoretical object to a phrase rather than to individual words. Let me give an example from some recent work I have been doing in another context. In a variety of computer applications and as a focal point of much research in computer science, there is a desire to develop question-answering systems so that when a question is input the computer can give back the correct answer. In analyzing a typical example much looked at because of its simplicity, namely, the geography of a set of countries, we might ask the question, "Does X have diplomatic relations with Y?" Now, if we take the simple approach that each single word designates a set-theoretical object, then the word relation in this context has a quite abstract set-theoretical object as its denotation. But, if instead, we take the phrase diplomatic relations as a denoting phrase, the parts of which do not denote, then a much simpler and more concrete set-theoretical object can be assigned to that phrase, namely, just what we expect as the ordinary binary relation between countries.

For some people the talk about set-theoretical objects will already seem somewhat abstract and perhaps obscure. It should be kept in mind that by set-theoretical object I ordinarily mean a fairly simple object like a class of individuals, a binary relation between individuals, etc. In ordinary talk anyway it is unusual to have set-theoretical objects of any really great complexity denoted by words or phrases occurring in the talk.

The next point of importance about the application of model-theoretic semantics to natural languages, as well as to formal languages, is that we cannot give an adequate account of meaning by assigning a denotation to individual words or phrases, or as we would tend to say in grammatical context, by assigning denotations to the terminal words. Set-theoretical functions now enter in a second way, namely, in telling us how denotations of the various parts of the sentence are related. The analysis of how the various parts of a sentence are related in terms of meaning, that is, to put it explicitly now, what set-theoretical functions relate the denotations of the words occurring in the sentence, constitutes one important part of our intuitive idea of meaning.

Again, as in the case of the denotations of individual words, the set-theoretical functions that relate the denotations of individual words are ordinarily relatively simple in character. If I use the phrase red books, for instance, then the natural set-theoretical function for this phrase is the intersection of the set of red things and the set of books. The subtle thing about the semantical functions relating the various parts of a sentence is that the surface evidence for the choice of these semantic functions is considerably less evident than is the choice of the denotations of individual words or phrases. As far as I can see there is no escaping this difficulty. In one genuine and obvious sense, the semantic functions that represent the structure of the meaning of a sentence are theoretical in character. The correctness of a given choice cannot be settled by any direct observational procedure, but rather only by indirect procedures of confirming predictions, as for example, confirming a variety of predictions about responses to questions, executing actions taken in response to commands, etc. On the other hand, using a weaker standard of introspection, in many cases the selection of a particular set-theoretical function, seems obvious and natural to any native speaker of the language. It seems to me, for example, that this is the case with the selection of intersection in the case of red books. However, an example already given shows that this selection of function will not work uniformly with adjectives, namely, in the phrase alleged dictators.

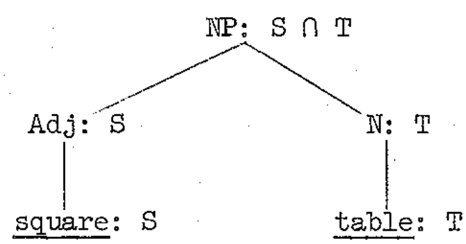
If we look at a sentence of any complexity, it is natural to ask how the semantic functions that express the structure of the sentence, that is, the relationships between the denotations of the individual sentences or phrases, are to be built up. Fortunately, a straightforward answer is available to this question. With each production rule of the grammar there is associated a semantic function, and thus, we may convert each derivation tree for a given terminal utterance to a semantic tree by attaching not only labels to the nodes of the tree, but also denotations generated by the semantic functions.

Let us illustrate these ideas with some simple examples. Consider first the rewrite rule

$$NP \rightarrow Adj + N .$$

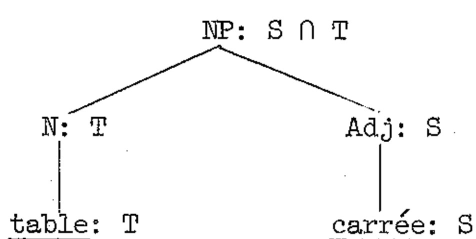
The simple semantic function associated with this production rule is intersection of sets, as already discussed above. Using this production rule, let us construct the semantic tree for the phrase square table. Let S be the set of square-shaped things and T the set of tables. The denotation of each node of the tree is shown after the colon following the label of the node.

(1)



The semantic tree for the corresponding phrase in French looks very similar, except that a left-right reflection is made; however, the denotation of NP is left undisturbed, because intersection of sets is commutative.

(2)



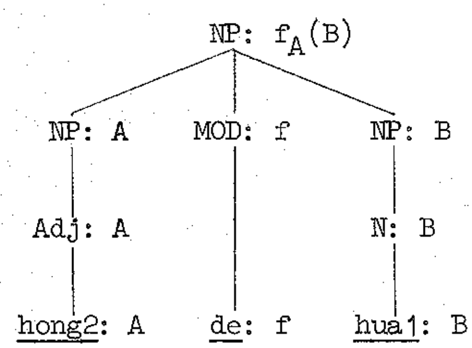
I would like to say that an analogous use of intersection as the semantic function attached to the generating rule for simple noun phrases will suffice in a wide variety of languages. However, it is doubtful that this is the case, mainly because the grammatical structure of noun phrases is different in other languages. Consider an example from our Chinese corpus, written in pinyin notation with tones indicated by the numerals 1-4 to make possible linear processing of computer input and output. The example is *hong2 de hua1*, literally, *red of flower*, and more idiomatically, *red flower*. Because of the extensive use of the particle *de* (or *te*), restraint seems required in classifying hong2 (*red* or *redness*) as an adjective. The semantical structure of this Chinese phrase is much like the English *capitol of France* or the French *capitale de la France*.

To draw the semantic tree of the Chinese phrase, we need some notation. Let MOD be an adjective-or-adverb-forming particle. For sets A, B and f, where f intuitively is a function, let ψ be the set-theoretical function defined as:

$$\psi(A,B,f) = f_A(B) ,$$

and f is a choice function such that for each A, $f_A(B) \subseteq B$.[2]

The tree looks like this.



Note that A is the set of red things, B is the set of flowers, and f is a function that selects the set of red flowers from the set of flowers. In other words, $f_A(B)$ is the set of red flowers.

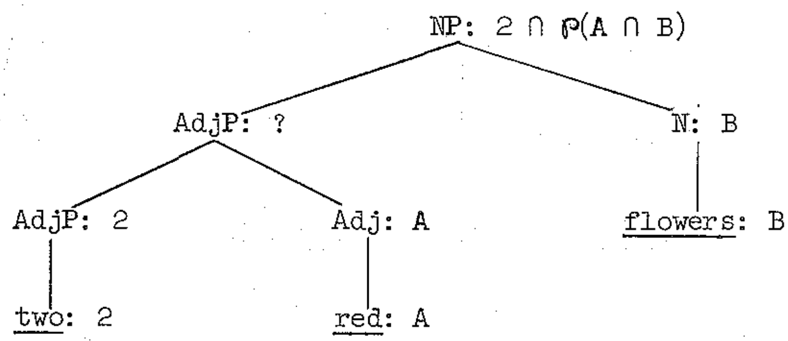
Numerical adjectives. That the model-theoretic semantics can determine a choice between the generative or production rules of the grammar constructed for a corpus is nicely illustrated by the simple attributive use of numerical or cardinal concepts in children's speech. Let us begin with English and then look at some corresponding French and Chinese phrases.

An example as good as any is two red flowers. A part of our noun-phrase grammar, very close to the one I developed earlier for Adam I in Suppes (1971), might look like the following:

$$
(3) \qquad
\begin{aligned}
NP &\rightarrow AdjP + N \\
AdjP &\rightarrow AdjP + Adj \\
AdjP &\rightarrow Car \\
AdjP &\rightarrow Adj\ .
\end{aligned}
$$

Here "AdjP" is a nonterminal symbol used to obtain a simple recursion for building up adjective phrases, "Car" is a nonterminal symbol for cardinal number names, and "Poss" is, of course, for possessives. The last three rules of this grammar would most naturally have the identity function as its semantic function: each set is mapped into itself, and in the simple case the first two rules would have set intersection as the appropriate semantic function. Both identity and intersection functions have been used already; in trees (1) and (2), the lexical rules replacing Adj by square, etc., have the identity function as the semantic function. The semantic tree for two red flowers according to the grammar (3) would look like this.

(4)



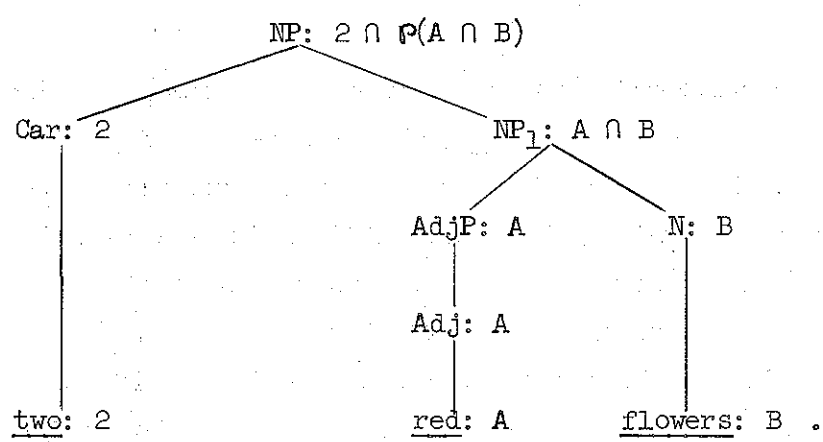
I have written ? for the denotation that is not assignable at the node labeled "AdjP". The notation for the denotation of the root of the tree may look formidable, but its intuitive meaning is simple. I use the Frege-Russell concept of cardinal number: 2 is just the set of all pair sets. (To avoid standard paradoxes of set theory, I only consider members of sets a certain distance up the hierarchy of sets, functions and relations--this is a technical problem of no real concern here.) The set

$A \cap B$ is, as before, just the set of red flowers, and $\wp(A \cap B)$ is the power set of $A \cap B$, i.e., the family of all subsets of $A \cap B$.

It is important to realize that I am not suggesting that a speaker or listener of English is examining in any sense the entire set 2 or the large set $\wp(A \cap B)$. A model of language is being provided within a standard set-theoretical framework. To provide a psychological theory of how the child comes to understand these denotations is a matter that in my judgment requires *still more* set-theoretical machinery, not a different sort of mathematical framework from the set-theoretical one I am using.

There is a simple solution to our problem of the grammar of *two red flowers*. It is to let the semantics guide the construction of the tree, and thus of the generative rules. The tree we want is something close to the following:

(5)



And the partial grammar (3) should be revised to:

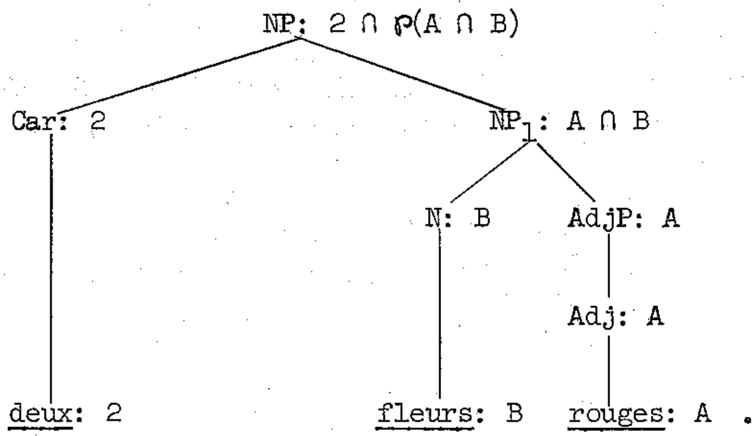
(6)

$$
\begin{aligned}
NP &\rightarrow Car + NP_1 \\
NP_1 &\rightarrow AdjP + N \\
AdjP &\rightarrow AdjP + Adj \\
AdjP &\rightarrow Adj\ .
\end{aligned}
$$

In (5) and (6) the subscript "1" on "NP" has been introduced to impose a restriction that blocks a recursion of cardinal number names. At any simple and straightforward level, we do not want phrases such as *two three red flowers*.

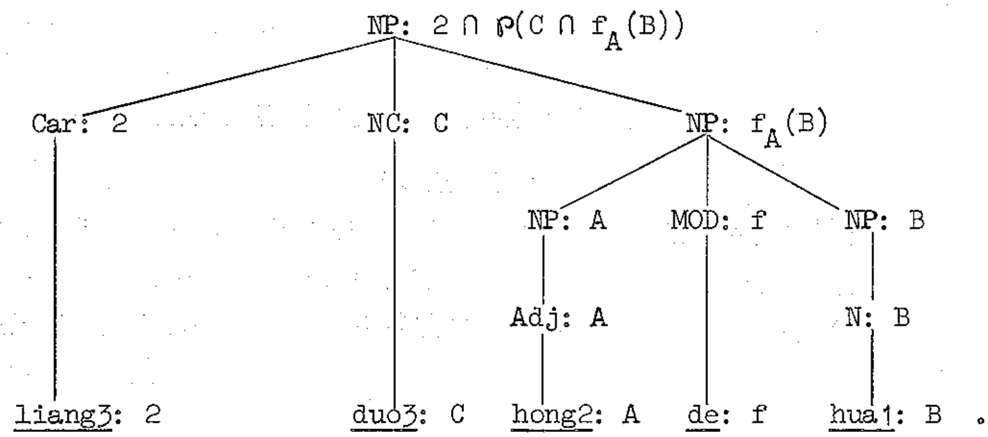
The French phrase corresponding to *two red flowers* is *deux fleurs rouges*, even though it is much more uncommon in French to omit the definite article than it is in English. The semantic tree is the same as (6), except for the sort of left-right reflection that occurred in going from (1) to (2):

(7)



The corresponding Chinese semantic tree that includes a noun classifier (NC) and the particle *de* (MOD) is more complicated on the surface than the English or French trees, but the underlying semantics is similar. (Later we shall look at some children's phrases in Chinese that omit the noun classifier or particle, thus making them closer in surface structure to the English or French examples.)

Teresa Cheng and I currently feel that the simplest semantics for the noun classifiers is to let them denote the union of all the sets of objects denoted by the nouns they modify. (When an NC is used as a mechanism of pronominal reference something more must be said.) On this assumption, our semantic tree for *liang3 duo3 hong2 de hua1* (two red flowers) is:



Sample data. To show that the semantic functions I have been discussing are to be found in children's speech under the natural interpretation of what they are saying, I give some examples from Nina (English), Philippe (French) and Chi-Chi (Chinese).[3] All three of the children are between two and three years of age, but Nina and Philippe were closer to two than three at the time the particular instances were recorded.

I begin with the intersection function for Adj + N. In the Chinese examples the particle de does not occur. Instances in which it does are listed below.

Intersection Function

red fish	aiguille rouge	hong2 hua1er1 (red flower)
big bird	grosse raquette	bai2 yi1shang (white dress)
big kitty-cat	bon côté	xiao3 yang2la4 (little candle)
big mousie	petits ronds	da4 gong1dian4 (big palace)
tiny rabbit	pauvres voitures	jin1 ji1dan4 (golden egg)
tiny guitar	petite aiguille	hei1 mian2yang2 (black lamb)

In the case of French the position of the adjective does not change the semantic function, but a preliminary scan of the corpus does show a greater frequency of adjectives before nouns than the reverse in the early recordings of Philippe. The somewhat greater sophistication of the Chinese examples is at least partly a reflection that Chi-Chi is about six months older than Nina and Philippe.

Next, let us look at the choice function as the semantics of possession.

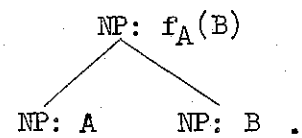
Choice Function for Possession

Mommy eyes	raquette de papa	mián2yang2 de mao2 (lamb's hair)
rabbit splinter	cordes de la raquette	ma1ma1 de hua4 (mother's talking)
horse feet	trains de tracteurs	Qi3qi3 de shu1 (Chi-Chi's book)
dolly dress	peau de Philippe	ge1ge1 de shu1 (brother's book)
		ba4ba4 de shu1 (father's book)

The same uninflected patterns of possession in English are exhibited in Adam I. Analysis in terms of a choice function is given in Suppes (1971) and will not be repeated here, except to note that for the English grammar, one production rule is

$$NP \rightarrow NP + NP$$

and the semantic tree is



Use of cardinal number names in noun phrases as already discussed above is illustrated in the following examples.

Number Function

one rabbit	une deux trous	wu3 ge4 xiao3tou1 (five thieves)
three ball		
two ladies		

In the Chinese example, the particle de does not occur, only the noun-classifier ge4. I emphasize that the frequency of cardinal number names is low in all three corpora.

Concluding Remarks. Within the confines of this paper I have restricted myself to some of the simplest examples of semantic functions in the speech of young children. I believe the more complete identification of the set of such functions, and especially their sequence of appearance in the language of the child, will provide a new and significant way of looking at language acquisition. The relation of such functions to the linguistic concept of "deep structure" is apparent, but for a number of reasons that I cannot develop here I am not inclined to identify the two approaches.

An equally important aspect of model-theoretic semantics in the analysis of children's language is comparison of semantic functions across languages. The examples I have given bring out semantic similarities of English, French and Chinese, but they are really meant only to exhibit the methodology. More detailed and more quantitative comparisons are needed to assess the similarities and differences in a serious way.

Finally, I reiterate the main purpose of this paper. It is to show in an informal way how model-theoretic semantics may be used to give a straightforward analysis of the meaning of children's language. Such an analysis is an essential element of any empirically adequate theory of language or language acquisition. That a systematic account of meaning is lacking in most discussions of language acquisition is surprising, at least on any common-sense view of what aspects of language are important. The methods I have outlined, which derive from the formal work of Frege in the 19th century and Tarski in the 1930's, can help to fill this lacuna.

References

Gammon, E. A syntactical analysis of some first-grade readers. Technical Report No. 155, June 22, 1970, Stanford University, Institute for Mathematical Studies in the Social Sciences. To appear in K. J. Hintikka, J. Moravcsik and P. Suppes (Eds.), *Approaches to Natural Language*, Reidel Publ. Co., 1972.

Suppes, P. Probabilistic grammars for natural languages. *Synthese*, 1970, *22*, 95-116.

Suppes, P. Semantics of context-free fragments of natural languages. Technical Report No. 171, March 30, 1971, Stanford University, Institute for Mathematical Studies in the Social Sciences. To appear in K. J. Hintikka, J. Moravcsik and P. Suppes (Eds.), *Approaches to Natural Language*, Reidel Publ. Co., 1972.

Footnotes

[1]The research reported here has been supported in part by the National Science Foundation under Grant NSFGJ-443X, and by the Office of Naval Research under Contract N00014-67-A-0112-0049.

[2]Such choice functions arise early in children's speech; I used them in Suppes (1971) in the analysis of Roger Brown's classic corpus Adam I. In the earlier article I required that $f_A(B) \in B$, which makes f a standard set-theoretical choice function. I have come to feel that the better choice is that $f_A(B) \in \mathcal{P}B$, i.e., in the power set of B, which is the set of all subsets of B, and for this purpose, we may write, as I have in the text, $f_A(B) \subseteq B$. In the present case, we end up with

$$f_A(B) = A \cap B \; ,$$

and it might be asked why not dispense with the function f and not let de denote at all? My present view of the matter is that we assign the semantic function ψ to the rule

$$NP \rightarrow NP + MOD + NP \; .$$

We may want to replace the first NP not by an adjective, or an adjective-like word, but by a noun expressing possession as in ge'ge' de shu' (brother's book). Then intersection is totally inappropriate. Here $f_A(B)$ is the set of members of B possessed by A.

[3]The recording and transcribing of Nina's speech has been done by Mrs. Florence Yager of the Institute's staff.