THE IDENTITY OF NOUN PHRASES

Lauri Karttunen

December 1967
THE IDENTITY OF NOUN PHRASES

Lauri Karttunen

December 1967

Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the policy of any of its governmental or private research sponsors. Papers are reproduced by The RAND Corporation as a courtesy to members of its staff.

This paper was prepared for presentation at the Forty-Second Annual Meeting of the Linguistic Society of America, December 28, 1967, Chicago, Illinois.
This paper attempts to discuss certain difficulties in dealing with noun phrases that plague the current theory of transformational grammars. By 'current theory' I mean essentially the kind of concept of grammar proposed by Noam Chomsky in *Aspects of the Theory of Syntax* (Cambridge, Massachusetts: The MIT Press, 1965). Thus I wish to contribute to the growing criticism of this approach by the neo-Chomskian generation: James McCawley, John Robert Ross, George Lakoff, et alia.

There seems to be a great uneasiness among the people actively engaged in theoretical work with the conservative transformational notions. Considering recent radical suggestions on how the theory should be changed to take care of one or another rediscovered fact about language one might wonder whether the term 'transformational generative grammar' be at all appropriate any more. While granting enough room for the semantic drift of such labels, it is worth pointing out that the new ideas that are now being circulated are mostly incompatible with and often opposite to the beliefs of early transformationalists. Such is the case with this paper.
Let us consider the problem of noun phrase identity. It has been assumed for a long time that many pronouns of the surface structure are derived from underlying more fully developed noun phrases by deletion. This is the case with the pronouns who, his, and himself in (1a)-(3a). The underlying structures of these sentences would be something akin to (1b)-(3b).

(1a) The boy who Mary loves kissed Jane.
(2a) John scratched his arm. ('his own arm').
(3a) The barber shaved himself.

(1b) The boy WH Mary loves the boy kissed Jane.
(2b) John scratched John's arm.
(3b) The barber shaved the barber.

Since deletions have to be, in some sense, recoverable, transformational theory requires identity between the deleted part and some part in the remaining sentence - or at least non-distinctness. It is further understood that deletions will turn out to be reducible to a few and possibly universal rules. In the above examples the deletions are attributed to such a putative universal, often called IDENTICAL NP DELETION. But it has proved difficult to specify what exactly constitutes the kind of identity under which deletions take place. In the case of noun phrase deletions, the following three criteria are usually mentioned.
Identical constituent structure
Morpheme-for-morpheme identity
Coreferentiality

There is also a fourth condition, namely the identity of inherent features, mentioned by Chomsky (Aspects, p. 182) that weakens the identity condition into a requirement of non-distinctness. However, the question which features are inherent and which are added by transformations has never been fully worked out.

It seems that originally only the two first criteria were assumed. The third condition, coreferentiality, became necessary when it was agreed upon that transformations should not change meaning. Thus it became no longer possible to derive e.g. (3a) from (3b) by an optional deletion rule. Although the identity of external reference has been repeatedly pointed out as being now a necessary condition for NP deletion, the real consequences of this innovation have perhaps not been fully understood, as we will see later in this paper.

There are some well-known problems with the universal notion of identity. I wish to demonstrate once again briefly that there is a group of cases in which the above set of three conditions is not sufficient and then turn to another group of cases in which they do not seem necessary.

Consider the examples (4a)-(6a).
(4a) John scratched his arm and so did Mary.
(5a) I saw a lion and so did you.
(6a) The barber shaved himself before I could.

(4a) is an example recently discussed by John Ross ("A Note on 'Command'", mimeographed paper, 1967). The second clause so did Mary could be derived either from (4b) or (4c).

(4b) Mary scratched her arm.
(4c) Mary scratched his arm.

In the case of (4b), the deletion takes place only if both of the clauses in question are reflexive; that is, if his and John refer to the same person and, at the same time, her and Mary refer to the same person. No deletion takes place if only one or neither sentence is reflexive. Notice especially that in this case there must be two different arms involved.

In (4c) the requirement seems to be that the arm scratched is the same. It does not matter whether his in (4a) and (4c) refers to John or not, but it seems doubtful that (4a) could be understood to mean 'John scratched his right arm and Mary scratched his left arm'. Although the case of (4c) thus is easily explained in terms of the conditions (i)-(iii), there is no simple explanation for the reflexive interpretation of (4a) that violates in part both the morpheme-for-morpheme identity and the coreferentiality condition. The fact that both clauses must be reflexive is not easily captured in terms
of the three traditional conditions.

Consider now the case of (5a) that for many speakers is interpretable either in the sense of (5b) or (5c) which in turn seem to be reductions of (5d).

(5b) I saw a lion and you saw it too.
(5c) I saw a lion and you saw one too.
(5d) I saw a lion and you saw a lion too.

In (5b) a specific, even if indefinite, lion was seen by me and you; in particular, the lion we saw was the same. In (5c) the lions seen may or may not have been the same, the sentence really does not suggest anything one way or another but only states that we both perceived one member of the set of lions, a "generic" lion so to speak. It would appear then that (5d) reduces to (5a) provided that either both of the NP's in question have a specific reference and the reference is the same (the sense of (5b)) or neither of them has a specific reference (the sense of (5c)).

Compare now sentences (4a), (5a), and (6a), which is one of Chomsky's examples at the 1966 Linguistic Institute. They all appear very similar. The deletions - if there indeed are any - presuppose some kind of semantic equivalence between the two clauses in question, and it just happens that this equivalence can be achieved in all of them in more than one way. All three are ambiguous, but not quite the same way. Unfortunately it also appears that this intuitive notion of semantic equivalence, which
seems to play a role here, is not related to the three traditional conditions in any simple fashion. It would have to be a part of a universal semantic theory. Needless to say, presently we are unable to suggest any definition for such a notion. But if there is a simple universal condition under which deletions take place, it must be a semantic notion. The alternative is a Boolean expression of awful complexity stated in terms of structural and morphemic identity, and of coreferentiality, which we cannot write either.

Let us now turn back to the more simple cases of (1)-(3).

It has been shown that, for pronominalization at least, coreferentiality is a necessary condition. Could it also be made a sufficient one? The answer to this depends partly on the question to which constituent the referential meaning is to be attached. Originally Chomsky proposed (Aspects, p. 145) that, by general convention, each occurrence of a referential item be assigned an integer as a feature. This would make the referential index a part of the complex symbol of nouns. The same view was held by Paul Postal in his article on pronouns ("On So-called Pronouns in English," 17th Annual Round Table, Georgetown University Press, Washington, D.C., 1966).

More recently, James McCawley has argued quite convincingly ("How to Find Semantic Universals in the
Event that There Are Any", mimeographed paper, 1967) that referentiality is a property of noun phrases rather than individual nouns.

To assume that McCawley is right and that the referential index is raised from the noun to the NP node actually renders the requirements of identical constituent structure and morpheme-for-morpheme identity meaningless. This comes from the fact, observed by Postal (above cit., p. 202) and others, that structures of the form of (7) cannot possibly be interpreted with the two NP's coreferential.

(7) The big boy saw the red-headed boy in the mirror.

In Postal's words: "It seems then that there is a principle of language which requires identical indices to occur only in nouns which have identical 'dominating constituents'". In his opinion identical indices on the two occurrences of the noun boy in (7) would presuppose the structural and morphemic identity of the two noun phrases the big boy and the red-headed boy. Since this is not the case, there must be two boys involved. As Postal quite correctly comments, it does not look as if this fact were a peculiarity of the grammar of English. It is probably true of all natural languages.

If the referential index belongs to the node NP, we simply have a situation where two noun phrases must necessarily be structurally and morphemically identical.
provided that they have identical referents; but the reverse does not hold, that is, structural and morphemic identity does not guarantee the same referent.

At least two ways are now open for the treatment of noun phrase identity. According to the first one - let us call it Proposal 1 - the general theory would contain a general well-formedness condition on phrase markers to the effect that constituents with the same referential index must dominate identical subtrees. This kind of condition is in principle unexpressable in terms of the usual phrase structure rules, even if viewed as tree-formation rules instead of conventional rewrite rules. Assuming that there is a choice of subrules, of two co-referential NP's, neither one could be expanded until the other has been rewritten. This problem has nothing to do with the question whether the rules of the base are ordered or not, it is simply a consequence of sequential derivation. The condition can only be stated, if the rules of the base are understood to be not generative rules at all in the usual sense but well-formedness conditions on trees, as McCawley has already argued on independent grounds ("Concerning the Base Component of a Transformational Grammar," mimeographed paper, 1966). In his view the rules of the base are not sequential rules that apply one at a time. Instead a certain set of WF-conditions as a whole is either fulfilled or not fulfilled. Another way to put this would be to say that WF-conditions
all apply simultaneously. In such a system it would be meaningless to ask whether it is, say, the subject NP that is prior and the coreferential object NP that is a copy or vice versa. The theory would allow for no such ordering; it would only specify that NP's with the same referent must be structurally and morphemically identical.

There is another way of dealing with coreferential noun phrases - call it Proposal 2. Suppose that for each set of NP's with the same referential index there be only one NP, say the highest one in the tree, that dominates a subtree with inserted lexical elements. The remaining coreferential NP's are terminal symbols associated only with their referential index. They receive an interpretation only after the antecedent NP has been evaluated semantically. We might, for example, have a universal agreement rule that distributes all the semantic properties of a given NP to all NP's which have the same referent. The underlying structure of (8) would then look something like (9).

(8) My buxom neighbor introduced herself.

(9)
The semantic rules of the grammar associate with the subject NP a set of properties, such as HUMAN, FEMININE, SINGULAR, etc. The above agreement rule assigns these features also to the object NP, which eventually becomes the pronoun herself in the surface structure.

According to the proposal 2 there would be no rule, such as IDENTICAL NP DELETION that is conditioned by structural and morphemic identity, coreferentiality, or the like. The structural descriptions of certain transformations, such as pronominalization, will specify noun phrase identity by requiring that the referential indices of the NP's in question be the same, but nothing need be deleted.

Both of the above proposals undoubtedly have inherent problems that may prove difficult to overcome. Let us for the time being assume that they can be made consistent and precise. Is there any empirical evidence that would help to decide (a) between the present theory and the new proposals, and (b) between proposal 1 and proposal 2? There seems to be at least some. In the current theory, the interesting observation that sentences of the type of (7) must refer to more than one person would appear to be only an accidentally true statement about languages thus far studied. It would not matter, if it would turn out to be false. On the other hand, both of the new proposals require that this be true for
all natural languages. Thus they restrict more narrowly
the notion of a possible natural language, and have to be
preferred in the absence of contrary evidence.

The problem of choice between proposal 1 and proposal
2 is a more intricate one. It depends partly on how one
deal with the ironical interpretation of sentences such
as (10a) and (11a).

(10a) Mary loves Mary.
(11a) John trusts nobody but John.

(10b) Mary loves herself.
(11b) John trusts nobody but himself.

Normally these are not understood as synonymous
with (10b) and (11b), but it is not difficult to find
instances in which people in fact do impose such as in-
terpretation on them - be it then for the sake of some
special effect, irony, or what not. How is this anomaly
described in terms of the grammar?

According to proposal 1, the grammar includes an
obligatory rule for deleting identical noun phrases.
Sentences such as (10a) and (11a) would be violations
to the deletion rule, hence no pronominalization. In
the second proposal there is no such deletion rule; in-
stead, the underlying representations of (10a) and (11a)
must contain different referential indices for the super-
ficially identical noun phrases. The situation is out-
lined in (12).
Proposal 2 makes the deviant interpretation of (10a) a semantic contradiction rather than a violation of any syntactic rule. The referential indices must be different but the referent is understood to be the same. (10a) would then belong to the large number of cases where the regular and the intended meaning clash; e.g. A big deal!, And Brutus is an honorable man.

It seems quite legitimate to ask under which conditions such deviant interpretations are actually possible. If they are derived by violating the obligatory deletion rule, the resulting deviant sentence should always contain two identical surface NP's. Consider now the following examples (13) and (14).

(13) Lyndon Johnson trusts Lyndon Johnson.
(14) The President trusts Lyndon Johnson.

Assuming that Lyndon Johnson indeed is President, the second sentence appears quite on par with the first one. The examples can only be understood as some kind of ironical remarks, and in this sense, synonymous with each other. The fact that (14) seems to deviate from
normal sentences exactly in the same manner as (13) cannot be explained in the framework of the first proposal, since the NP's in question are not identical in (14).
Proposal 2, on the other hand, labels (13) and (14) as identical semantical contradictions. It would then be expected that they be interpretable in the same manner, if interpretable at all. This seems to be the case.

This kind of empirical evidence is admittedly of extremely marginal kind, but it is difficult to do better at this point. Let us tentatively adopt the second proposal and rephrase it as follows.

(15) In the deep structure, some NP's are marked as referential by assigning an index to them. For each set of NP's with identical indices, only the topmost NP is expanded to a subtree in which lexical items are inserted. The rest of the coreferential NP's are unexpanded terminal symbols. The semantic component of the grammar determines the meaning of the expanded NP by processing the subtree it dominates, and the resulting set of semantic features is assigned to all NP's with that particular referential index.

It should be obvious that (15) makes a very strong claim about the nature of natural languages. It is also in many ways incompatible with the notion of deep structure as Chomsky defines it in Aspects. The output of the base would no longer be input to independent semantic and transformational components, instead there has to be a semantic interpretation before any of the syntactic rules can operate. Chomsky's deep structure also is the level on which the co-occurrence restrictions on lexical
items are expressed. But if, for example, the underlying structure of (8) is something like (9), it is not possible to decide whether the restrictions on the verb to introduce are met until the unexpanded NP has been associated with a set of semantic features. All evidence we can find for (15) thus will speak directly against the deep structure as an independent level of linguistic description and contribute to the recent criticism by McCawley, Lakoff, and others.

Secondly, (15) restricts severely the class of possible underlying structures. In addition to the restrictions discussed above, it also excludes a certain kind of recursion, namely the unbounded nesting of coreferential noun phrases. This kind of nesting seems to have been allowed in all previous formulations of constituent structure. It would be the source of sentences of the form (16) which would have (17) as its underlying P-marker.

(16) The boy ( who ( who ( who is big ) kissed the girl ) ate the cake ) sang a happy song.
In (17) the part below the broken line is a violation of (15). Since the subject NP of the clause the boy ate the cake has the same referential index as another NP higher in the phrase marker, it should be unexpanded. Taken in isolation, the lower part of the P-marker (17) would naturally yield a grammatical sentence; it is only the recursiveness of coreferential embeddings that is ruled out by (15). Notice that sentences such as (18a) that might at the first glance appear to have an underlying structure of the form of (17) in fact do not. That is, the modifiers of the subject NP in (18a) are understood as forming a coordinate construction that is synonymous with the two conjoined relative clauses in (18b).

(18a) The big boy who kissed the girl ate the cake.
(18b) The boy who is big and who kissed the girl ate the cake.

Another case in which the set of underlying structures permitted by (15) is restricted in an interesting way, is illustrated by strings of conjoined adjectives, such as her wet wild beautiful hair. The fact that strings like this one are perceived as unstructured is difficult to explain in a theory that allows for structures such as (17). If the order of the adjectives is taken to reflect their position in the deep structure, the meaning should change when the order of attributes is changed; in general, it does not. If the surface ordering is free, or determined by some special properties of the adjectives in question, it should still be the case that strings of attributes are perceived as potentially ambiguous and the ambiguity should increase in proportion to the number of adjectives that are conjoined. This prediction also seems quite counterfactual. On the other hand, (15) only permits underlying structures of the type of (19) in which prenominal modifiers are derived from coordinate clauses.

(19) her wild wet beautiful hair

It is therefore crucial to show that cases, such as (20), (21), and (22) that have been presented as examples
of nesting adjectives, in fact are not derived from underlying structures of the type (17).

(20) A small dark room.
(21) The short happy life of Francis Macomber.
(22) His first important act.

These three sentences are all ambiguous in the same fashion. In one interpretation, the adjectives form a coordinate construction; for example, (20) refers to a room that is small and dark. The problem lies with the other interpretations: 'a dark-room that is small', 'the happy period, which remained short, in the life of Francis Macomber'. At the first glance these indeed seem to violate condition (15), but there are some good reasons for believing otherwise.

In a recent article ("Adjectives in English," Lingua, Vol. 18, No. 1 (1967)), Dwight Bolinger distinguished between two kinds of adjective modification. To put it simply, in one case, which Bolinger calls referent modification, the adjective tells us something about the intended referent; in the other case, called reference modification, the adjective goes with the head noun of the noun phrase and is not directly related to the intended referent. This distinction is illustrated by examples (23) and (24). (23) is a case of referent modification; it is the person Henry who is sleepy and who also happens to be a policeman. In (24) there is a subclass of police-
men called rural policemen to which Henry belongs, but the adjective rural has nothing to do with Henry as a person.

(23) Henry is a drowsy policeman.

(24) Henry is a rural policeman.

It appears that what is happening in examples (20) through (22) involves Bolinger's distinction between referent and reference modification. In (20) the adjective dark can be taken in either sense. It is either the case that both small and dark modify the referent or that small goes with the referent and dark with the noun room. Notice that in the latter sense it would not be contradictory to say something like (25).

(25) The small dark room is brightly lighted.

A similar argument could be given for (21) and (22). If the grammar incorporates a mechanism for accounting for the semantic difference between (23) and (24), it will than also provide for the ambiguity of (20) through (22). How this should be done in detail is not clear, but the problem is certainly not solved by underlying structures of the type (17).

By not permitting the nesting of coreferential noun phrases (15) makes the strong claim that such structures are not found in natural languages. Since this is quite a non-trivial restriction, the fact that it appears, by and large, compatible with reality is indirect evidence
Thus far we have only discussed condition (15) in connection with relative and reflexive sentences. The way (15) is formulated would suggest that it have a wider application. Unfortunately this turns out not to be the case. After trying to demonstrate the advantages of (15) over the present treatment of noun phrase identity, let us then point out where it fails. Consider sentences of the type (26) through (29).

(26) Mary saw the boy who she loves.
(27) It surprised John that he was becoming a father.
(28) She saw the boy who Mary loves.
(29) It surprised him that John was becoming a father.

Many English speakers will agree in saying that it is possible to interpret (26) but not (28) with Mary and she coreferential. In (28) there has to be two girls involved. There is a parallel relation between (27) and (29). This is exactly what we would predict from (15) which requires that the coreferential NP's lower in the tree become pronouns. For example, it should not be possible for the him of (29) to have John as its antecedent, since the latter is located lower in the base structure. But take now the passive of (28) and the unextraposed version of (29):

(30) The boy who Mary loves was seen by her.
(31) That John was becoming a father surprised him.

The consensus of English speakers is that (30) and (31) are exactly like (26) and (27), that is, it is natural to interpret them with the proper name and the pronoun referring to the same person. According to (15) this should not be the case. The proposal restricts too heavily the introduction of proper names and common nouns in dependent clauses. The worst possible case is illustrated by examples (32a) through (32d) in which there seems to be a complete freedom of whether the pronominalization takes place and of the order of pronominalization. It is possible to interpret them all as synonymous.

(32a) John, who Mary loves, kissed Mary.
(32b) John, who she loves, kissed Mary.
(32c) John, who Mary loves, kissed her.
(32d) John, who she loves, kissed her.

For the time being it is not clear what should be done about this. The failure of the present proposal in these last cases is no more spectacular than that of any other approach previously suggested. What is needed is more insight into the role played by the referential indices themselves. Except the work done by McCawley on plurals ("How to Find Semantic Universals.."), hardly anything has been written about the problem of reference in transformational grammars.

The question to be faced is this: What do referential indices refer to? The position tacitly assumed by Chomsky
in *Aspects* and generally followed elsewhere, including this paper, implies that the indices point to things in the real world as it is perceived by the speaker/hearer. There is also a more traditional view that distinguishes between two kinds of reference, one anaphoric, the other extralinguistic. The idea of the anaphoric use of pronouns is in fact imprinted in the term itself. In traditional terms, a pronoun has an antecedent which may be somewhere in the same sentence or somewhere in the previous discourse.

In retrospect it seems, that the underlying motivation for postulating a transformation such as IDENTICAL NP DELETION in early transformational studies was to capture the notion of anaphoric reference without actually talking about semantics. The failure of this policy seems pretty clear now.

I suggest then that we restore some of the traditional machinery and admit that there are two kinds of referential indices, call them anaphoric and demonstrative. The referential indices of pronouns (which up until now have been discussed under the pseudonym "coreferential NP" in this paper) and certain other noun phrases need not refer to the world but they may instead point to other nodes that reside either in the same sentence or somewhere in the past discourse. This would give our pursuit a more traditional flavor, but the likelihood of success with certain outstanding problems seems far greater.