

REFERENTIAL INDICES AS SYNTACTIC FEATURES: ANAPHORIC AGREEMENT IN GPSG*

Byung-Soo Park

This paper attempts to provide a GPSG account for anaphoric agreement in English, a phenomenon which used to be dealt with by pronominalization rules in classical transformational grammar. A new feature INDEX is proposed. It is a category-valued FOOT feature obeying the Foot Feature Principle. Grammatical agreement between pronouns and their antecedents is seen as a consequence of the distribution of the feature INDEX in local trees. The distribution is regulated by the interactions of the Foot Feature Principle and a slightly extended version of the Control Agreement Principle together with relevant Immediate Dominance Rules. Two particular assumptions need to be made for this account: direct (or indirect) objects control *that*-clause or *when*-clause complements and sentence-initial adverbial clauses are taken to be topicalized.

1. Introduction

The classical transformational approach to pronominal anaphors initiated by the pioneering works of Klima (1964), Langacker (1969) and Ross (1969) has long been abandoned.¹ Since then, a variety of interpretive approaches have been advanced. Details aside, these approaches assume that pronouns are generated in the base (instead of being derived transformationally) and that base-generated pronouns are interpreted semantically in the semantic component. The interpretive approaches regard conditions for anaphoric reference as a semantic problem, rather than a syntactic one. The recent development in the GB framework under the name of 'binding theory' seems to retain essentially the same view.

In Generalized Phrase Structure Grammar (GPSG) (cf. Gazdar, Klein, Pullum and Sag (1985) (GKPS, hereafter)), there is only a single level of representation for the syntactic structure of a sentence; there is no place for the distinction between deep and derived structures. Being a non-transformational theory of

*A preliminary version of this paper was read at the Language Research Seminar of the Language Research Institute of Seoul National University in December, 1985. I would like to thank Ik-Hwan Lee for his comments at the Seminar, and Chungmin Lee and Edward Keenan for their encouragement after the Seminar. I am especially grateful to Geoff Pullum, who read the draft of this paper and raised a question. My response to the question led me to write a new section, i.e., Section 4. I have profited both in content and in style from the critical comments of an anonymous reader of this journal, which also led me to add Section 4. Needless to say, none are responsible for my errors.

¹ One of the chief reasons for that was perhaps its inability to cope with the so-called Bach paradox arising from the following type of sentences: *The pilot who shot it down misidentified the plane he fired on.* This problem will not be dealt with in this paper.

grammar, GPSG cannot countenance a pronominalization transformation which would change full noun phrases into pronouns. In this sense, GPSG has something in common with the interpretive approach. However, such a similarity cannot be stretched any further. GPSG leaves it an open question whether defining conditions for anaphoric reference is entirely a semantic problem. Despite some discussion of reflexive and reciprocal pronouns, GKPS (1985) put aside the problem of ordinary anaphoric pronouns.

This paper addresses itself with this matter and attempts to show that it is possible to describe conditions for the interpretation of pronominal anaphors in terms of syntactic constraints. It will be shown that the description of those syntactic constraints is rendered possible by the theoretical advantages of GPSG, by its theory of syntactic features in particular. Specifically, this paper is concerned with the phenomenon of agreement between pronominal anaphors and their antecedents in a sentence.²

2. The FOOT Feature INDEX

As a starting point, I propose that anaphoric pronouns have a syntactic feature, INDEX. This is assumed to be a category-valued feature, a feature which takes a category as its value.³ For example, the personal pronoun 'she' should be represented as follows:

- (1) {<N, +>, <V, ->, <BAR, O>, <MALE, ->, <PLU, ->, <PER, 3>, <INDEX, NP>}

Note that the feature name INDEX has an NP as its value. Now the question is what kind of NP it should take as its value. Our discussion will be focused on this question shortly. Before doing so, however, it is necessary to clarify other assumptions.

I assume that INDEX is a Foot feature, obeying the Foot Feature Principle.⁴ The Foot Feature Principle may be stated informally as follows:

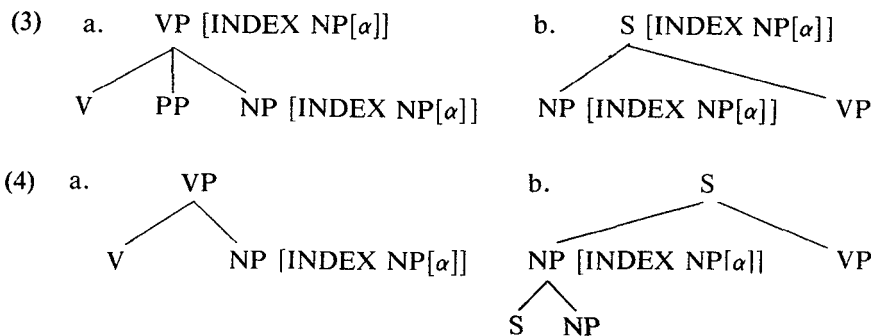
- (2) If a FOOT feature is instantiated on a daughter category in a local tree, it must also be instantiated on the mother category in that tree.

² Traditionally, grammatical agreement is an agreement between verbal and nominal elements (or between functions and arguments in logical terms), and is usually associated with some affixes attached to the verbal elements. Anaphoric agreement is not an instance of agreement of this sort in that it involves a relation between two nominals (i.e., between nominals and their antecedents). Nevertheless, many linguists tend to regard pronominal-antecedent relations as grammatical agreement. Pullum (1985) includes anaphoric agreement in his list of grammatical agreement (though he excludes it from discussion).

³ For the theory of syntactic features in GPSG, see Chapter 2, GKPS (1985). A syntactic category is defined as a feature bundle, not as a monadic element without internal structure.

⁴ Sells (1984) examined possibilities of treating pronouns using a FOOT feature and rejected them all. However, he did not examine the possibility of viewing INDEX as a FOOT feature.

For instance, trees such as in (3) are all admissible trees, which meet the Foot Feature Principle (FFP, hereafter), while those in (4) are not. (Suppose that there is no rule stipulating INDEX features in each case.)



In (3a), the FOOT feature INDEX is instantiated on the NP node, and therefore the same FOOT feature is also instantiated on the mother in the local tree. In (3b) once INDEX is instantiated on a daughter, it is consistently instantiated on the mother. However, (4a) violates the FFP because there is an instantiated INDEX feature specification on the daughter but no corresponding INDEX feature on the mother. For a similar reason, (4b) also violates the FFP.

The consequence of my proposal about INDEX is simply to add one more member to the set of FOOT features defined by GKPS (1985) as (5). We now have (6) instead.

(5) FOOT = {SLASH, WH, RE}

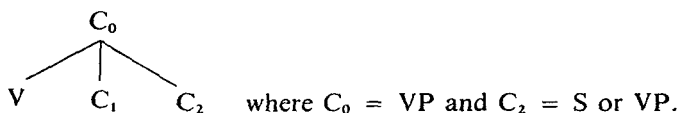
(6) FOOT = {SLASH, WH, RE, INDEX}

The value of INDEX is determined basically in the same way as other control features, i.e., by the Control Agreement Principle (CAP, hereafter). In GKPS (1985), there are two control features: AGR and SLASH. Now we have three, recognizing INDEX as a control feature as well as a FOOT feature. But the original CAP formulated in GKPS (1985) needs to be extended slightly in order to deal with anaphoric agreement as well as other agreement phenomena.⁵ The following provision should be added to the original CAP:

⁵ The original CAP may be phrased roughly as: (i) the feature specifications associated with the control feature (AGR or SLASH) of a controllee must be identical to those of its controller; (ii) if there is no controller, the feature specifications of the controllee must be identical to the specifications of the control feature of the mother in the tree. See Chapter 5, GKPS (1985) for its explication.

(7) Control Agreement Principle (Addenda)

In a local tree of the form below



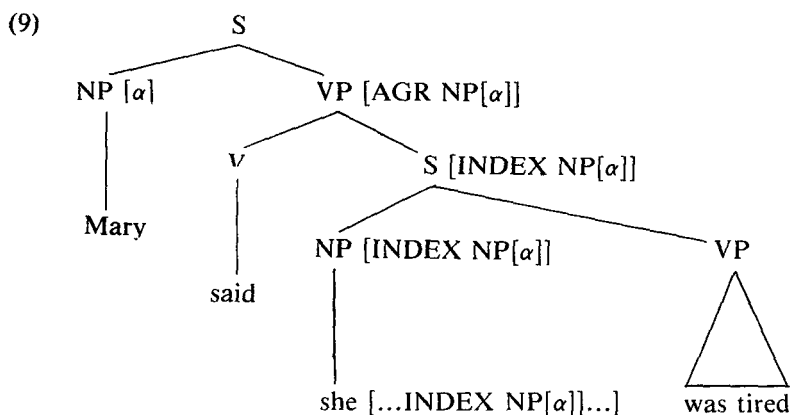
- (a) if C_1 controls C_2 , then the value of INDEX in C_2 is the same as C_1 .
 (b) if there is no controller, the value of INDEX in C_2 is the same as the value of AGR of C_0 .

3. Pronouns in *that*-clauses

Consider the following pair of sentences:

- (8) a. *Mary* said *she* was tired.
 b. **She* said *Mary* was tired.

The contrast may be accounted for correctly by the CAP (Addenda).



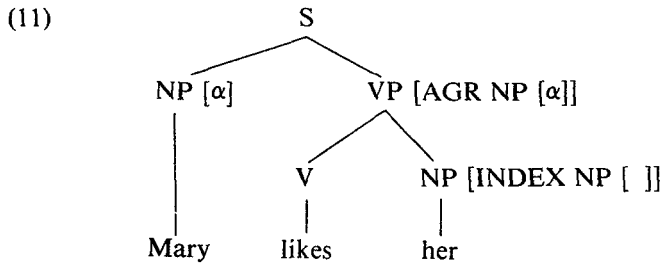
(NP[α] is an abbreviation of NP [3PER, -PLU, -MALE], more strictly, {<N, +>, <V, ->, <BAR, 2>, <PER, 3>, <PLU, ->, <MALE, ->}.)

The FOOT feature INDEX, which is part of the syntactic information included in the lexical item *she*, is percolated up to S in accordance with the FFP. Then the CAP (Addenda) operates to determine the value of the INDEX feature. In the local tree with VP and its two daughters V and S, there is no controller, and so the CAP (Addenda) (b) applies. It requires that the value of INDEX

be the same as the value of the AGR feature of the mother VP. On the other hand, the value of the AGR feature is the same as the subject NP (i.e., 'Mary') in accord with the original CAP as usual. Hence the value of INDEX has been determined properly, and both the FFP and the CAP are satisfied as they should be. It follows that the tree (9) is well-formed. In this way, the referential connection between the pronoun and its antecedent *Mary* can be correctly predicted.

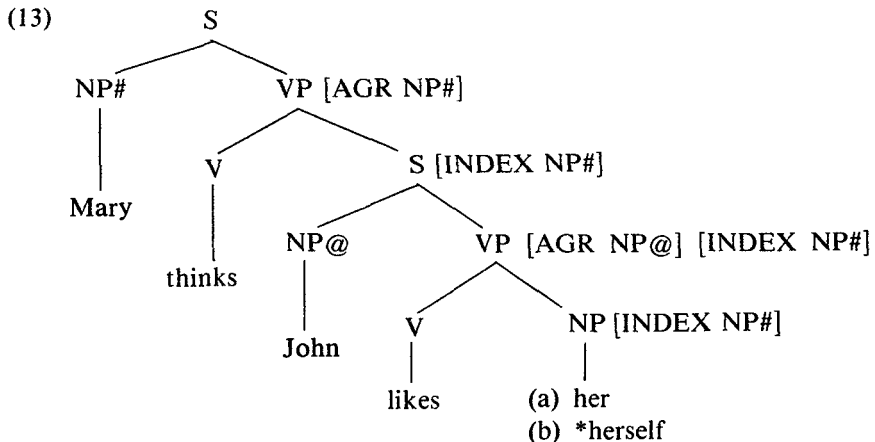
The condition that C_2 must be either S or VP in (7) above is necessary because of the fact that a noun which is a clause-mate of the subject in a simplex sentence is reflexivized, rather than being pronominalized. Observe the following pair of sentences:

- (10) *Mary* likes (a) *her.
 (b) herself.



This tree does not meet the condition because the constituent dominating the pronoun is neither S nor VP, and the CAP cannot apply to the tree; hence it is an inadmissible tree. The condition also accounts for the following contrast:

- (12) *Mary* thinks John likes (a) her.
 (b) *herself.

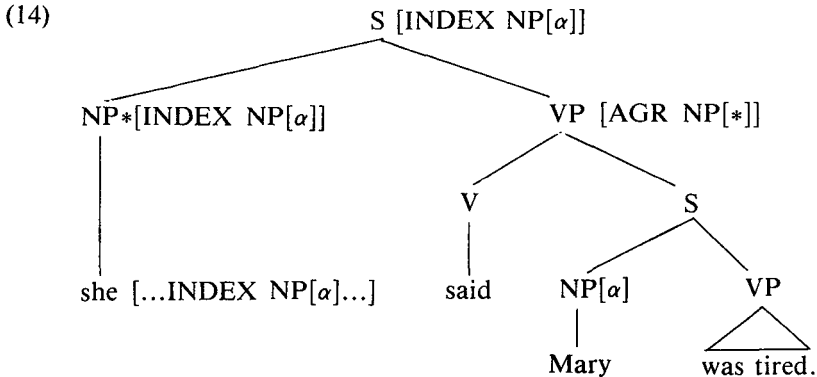


The pronoun *her* may be coreferential with *Mary*, not with *John*, while the reflexive pronoun *herself* with *Mary*, not with *Kim*.

The INDEX feature cannot receive its value from VP [AGR NP@] at the NP node dominating the pronoun *her* since the local tree configuration does not meet the condition for the CAP that C_2 must be either S or VP. However, if the INDEX feature passes up to embedded S node, then it is in a correct position to receive its value from VP [AGR NP#]

In the case of (12b), the REFL feature, which is a FOOT feature every reflexive pronoun has inherently, could not pass through the VP [AGR NP@] node because this VP is a generalized predicative in the sense of Pollard and Sag (1983).⁶ Hence the reference of *herself* should be determined within the boundary of the VP constituent. The sentence (12b) violates this constraint.

Now consider (14), the tree representation of (8b).



In (14), the INDEX feature of the pronoun *she* is percolated through NP, finally to the top S. This is of course guaranteed by the FFP. But there is no way of determining the value of the INDEX feature. None of the categories containing the INDEX feature is controlled by a category that would determine the INDEX value, and the CAP cannot apply. Consequently, the INDEX value remains 'unagreed,' so to speak. This means that the pronoun *she* does not agree with anything within the sentence. It follows that the subject pronoun *she* of the main clause and the subject noun of the embedded clause *Mary* cannot be coreferential.

⁶ What they call a Principle of R(efering)-pronoun Distribution is stated informally as: 'R-features cannot be adjoined to categories of generalized predicative type,' where a generalized predicative is a VP of type <NP,NP> or <NP,S>. This principle prohibits the REFL feature to pass through the VP [AGR NP@] node in (13). So a generalized predicative serves as a check point beyond which REFL cannot pass. Now INDEX is unlike REFL in that it can penetrate generalized predicatives, or any category for that matter.

In terms of tree admissibility condition, we say that the tree (14) is not admitted because a principle which must be applied cannot be.⁷ The principle in question in this case is the CAP.

In closing this section, a comment on the status of (8b) is in order. Note that a sentence like (8b) will be perfectly grammatical if the interpretation of the pronoun 'she' is free, not bound by another name within the sentence. In this case, the pronoun should be understood as a name rather than as an anaphoric pronoun, just as ordinary proper nouns like 'Kim' or 'Bill' are. It is contended, therefore, that non-anaphoric pronouns do not contain an INDEX feature specification, their interpretation having nothing to do with the FFP or the CAP. In other words, there are two kinds of English pronouns, anaphoric and non-anaphoric, only the former having an INDEX feature specification.

4. Two or More Indices in a Single Clause

In all the example sentences that we have considered so far, there was only one index feature in each case; in other words, there was only one anaphoric pronoun in a single clause agreeing with its antecedent outside the clause. Needless to say, there is no such constraint in English (or in any other languages). In fact, a clause may contain any number of pronouns whose antecedents lie outside the clause. For example, the complement clause in (15) contains two pronouns with antecedents outside the clause.

(15) John tells Mary (that) *she* should write to *him*.

Hence a constituent containing two pronouns should have two different INDICES, each INDEX taking a distinct value. Thus in (14), one INDEX feature associated with *she* should take NP *Mary* as its value and another INDEX feature associated with *him* should take NP *John* as its value.

Nevertheless, this is not permitted in GKPS (1985: 81) as it stands because of the self-imposed restriction which prohibits two or more FOOT features from appearing in a single constituent. The motivation for this restriction comes from the fact that English usually permits only one gap per constituent. As is conceded in GKPS, however, in order to cope with languages such as the Scandinavian languages in which double gaps are commonplace, this restriction must be eliminated or at least modified in some way. Even some English constructions make such a modification inevitable:

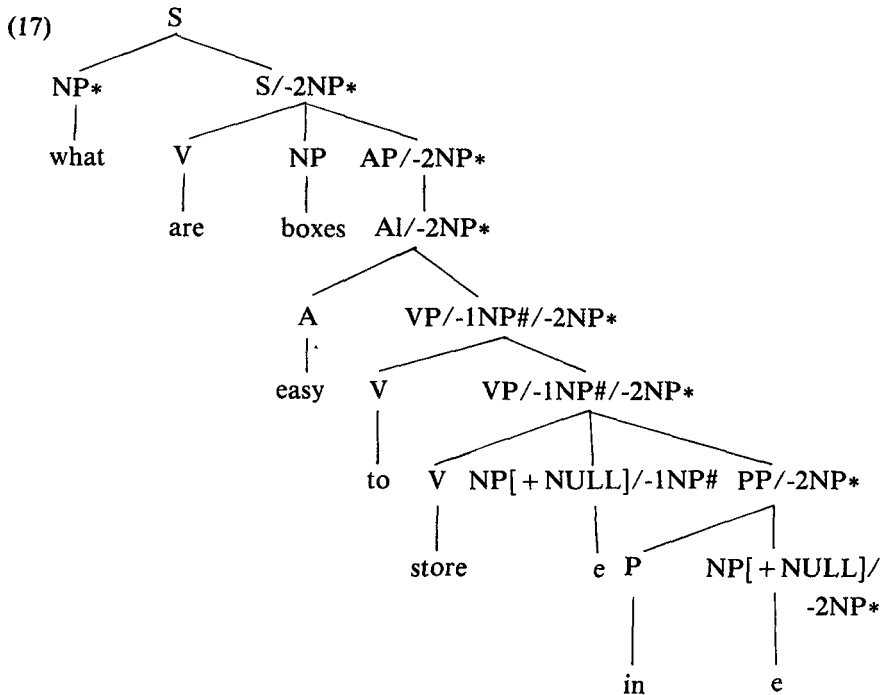
- (16) a. What are boxes easy to store _____ in _____?
 b. Which computer is this program likely to be easy to execute _____
 on _____?
 c. Someone that obnoxious, John would be difficult even for me to persuade _____ to invite _____.

⁷ For the notion of tree admissibility, see Section 5 or Chapter 5, GKPS (1985).

In each case, the infinitive complement in the *tough* construction has two gaps with distinct fillers outside the complement clause. In (16a), for example, the first gap will be filled by *boxes* and the second gap by *What*. This would mean that the complement clause should have two independent SLASHes.

Based on the facts about double gaps in sentences like (16) as well as in the Scandinavian cases, I suggest that the restriction against allowing more than two FOOT features in a constituent simply be dropped.⁸

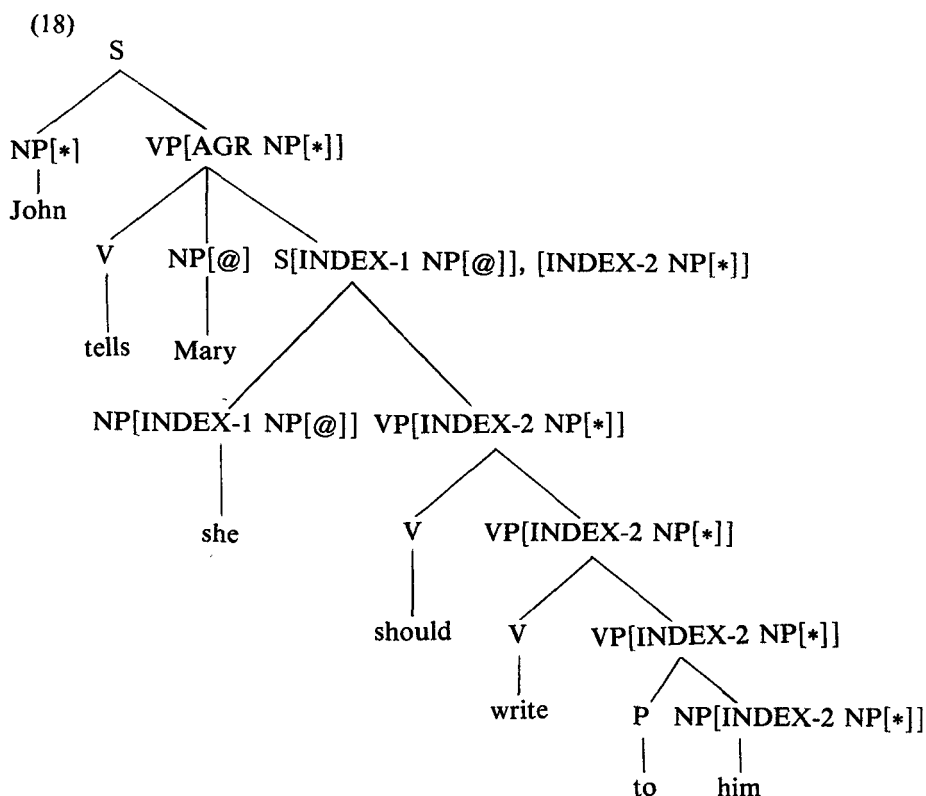
When more than two FOOT features appear in a constituent, we will need a device to identify each of them. This can be accomplished by assigning a distinct index to each. Then we would have SLASH-1, SLASH-2, SLASH-3, etc. at our disposal instead of having only one SLASH. Under this analysis, the structure of (16a) may be represented as follows (omitting irrelevant details):



⁸Gunji (1984) employs double gaps in a constituent freely when he describes interactions of topicalization and relativization and reflexivization as well as non-canonical word order variations in Japanese.

Note that the VP constituent dominated by A1 contains two SLASHes, SLASH-1 and SLASH-2. SLASH-1 is an inherited FOOT feature licensed by the Immediate Dominance Rule $A1 \rightarrow H[42], V2[INF]/NP[-NOM]$ (Cf. GKPS (1985: 150)), and for this reason, it is not instantiated anywhere above the VP. In contrast, SLASH-2 is an instantiated FOOT feature, and is instantiated throughout the tree.

Once the restriction imposed on the number of occurrence of a FOOT feature in a constituent is removed, sentences like (14) pose no problem as far as INDEX features are concerned. A single constituent is now allowed to contain as many INDICES as the number of pronouns occurring in it. Then each INDEX can be assigned a value as usual: i.e., in the same manner as in the cases where only a single INDEX appears in a constituent. Returning to (15), we can represent its structure as in (18) (again omitting irrelevant details).



Here we have two INDICES in the complement clause, and INDEX-1 takes its value from the controller of the clause, i.e. NP *Mary*⁹ while INDEX-2

⁹For some discussion of the notion of 'control', see Section 5.

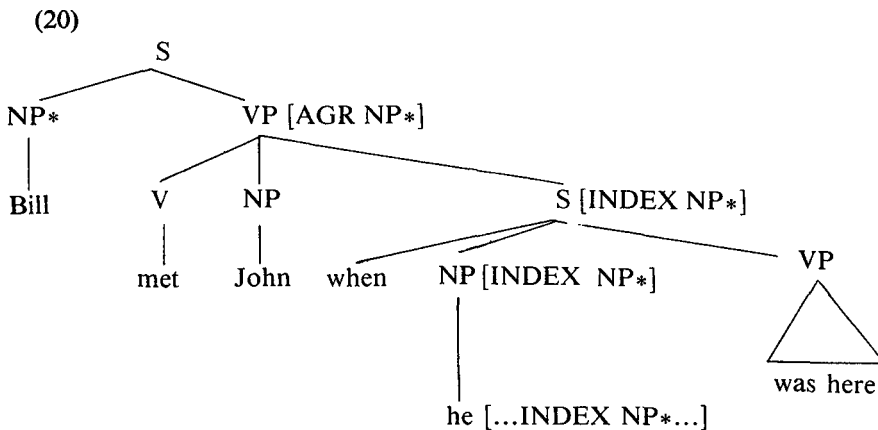
from the AGR feature of the VP in the main clause. The possibility that the two INDEX features take values the other way round would be blocked by gender conflicts in the corresponding NP categories. In case gender differences are neutralized (e.g., "John tells Bill he should write to him."), the possibility could not be blocked, and consequently *he* and *him* would be ambiguous syntactically. It seems that such an ambiguity should be resolved semantically, but I cannot discuss this problem any further here since the semantics of pronominal reference is beyond the scope of this paper.

5. Pronouns in Adverbial Clauses

Next we will consider cases in which embedded clauses are adverbial clauses (time, place, reason, concession, and the like), sentences like the following:

- (19) Bill met John when *he* was here.

This sentence is ambiguous; the pronoun *he* may be interpreted as coreferential either with *Bill* or with *John*. The case in which it is coreferential with *Bill* can be dealt with in exactly the same way as (9), where a *that*-clause occurs instead of a *when*-clause, assuming the correctness of (20) below for the structure of (19). Note, in particular, that the time adverbial clause is immediately dominated by the VP of the main clause.



The second reading of (19) appears to be problematic. The question is whether the direct object of the main clause controls the time adverbial clause. If we can answer this question affirmatively, we will have no difficulty accounting for the second reading. Once it is established that the direct object is the controller, with the time adverbial as its controllee, the FFP and the CAP will be

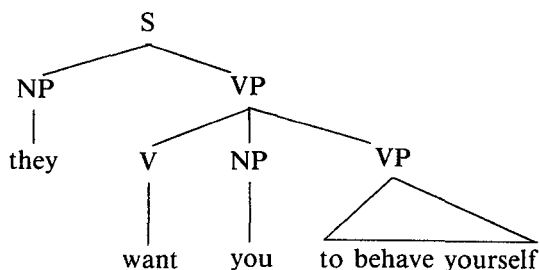
able to operate as usual, just as in all the previous cases.

GKPS (1985: 87) give an informal definition of CONTROL: '... a category C is controlled by another category C' in a constituent C^0 if one of the following situations obtains at a semantic level: either C is a functor that applies to C' to yield a C^0 , or else there is a control mediator C'' which combines with C and C' in that order to yield a C^0 .'

As examples for the first case, consider S and NP constituents. A subject NP controls a predicate VP in an S constituent since the VP is a functor which applies to the NP to yield an S . In other words, an expression of VP is a function from NP -type denotations to S -type denotations, and then the function VP is controlled by the argument NP . The same relationship holds between a Det and an N in the NP constituent, i.e., the former is controlled by the latter, the former being a function from N -type denotations to NP -type denotations.¹⁰

The second condition for control involves the notion of control mediator. Control mediators are a class of verbs whose lexical type is $\langle VP, \langle NP, VP \rangle \rangle$, functions which combine consecutively with a VP complement and an NP object to yield a VP expression. Typical verbs of this type include 'raising-to-object' verbs (e.g., *believe*, *want*, etc.) and 'object-controlled-equi' verbs (e.g., *persuade*, *force*, etc.).

(21) They want you to behave yourself.



Since the semantic type of the verb 'want' is $\langle VP, \langle NP, VP \rangle \rangle$, i.e., in (21), *want* is a function from VP (*to behave yourself*) to a function from NP (*you*) to VP (*want you to behave yourself*), the NP object controls the *to*-infinitive VP . Then the CAP can account for the agreement between *you* and *yourself*.

I now would like to propose that an analogous situation obtains in structures like (19) above: the direct object NP *John* controls the time adverbial clause *when he was here*. This means that a verb like *meet* is a control mediator if the VP dominating the verb dominates an adverbial clause. The plausibility of this proposal, I believe, comes from the interpretation of a sentence like the following:

¹⁰ For the theory of semantic types, see Chapters 9 and 10. GKPS (1985).

- (22) a. I told *Sandy* that *he* should leave.
 b. I told *Sandy* to leave.

In the semantic analysis of (22b), *Sandy* is the controller and the VP *to leave* is its controllee because the semantic type of the verb *tell* is $\langle VP, \langle NP, VP \rangle \rangle$. On the basis of this, the fact that the subject of *to leave* is understood as *Sandy* is explained. In a similar way to this usual practice of GPSG, it seems to me, the fact that *Sandy* and *he* may be coreferential can also be explained. Suppose that the indirect object NP *Sandy* controls the *that*-clause. Then, according to the CAP, the value of the INDEX feature of the pronoun *she* will be filled by the indirect object NP.

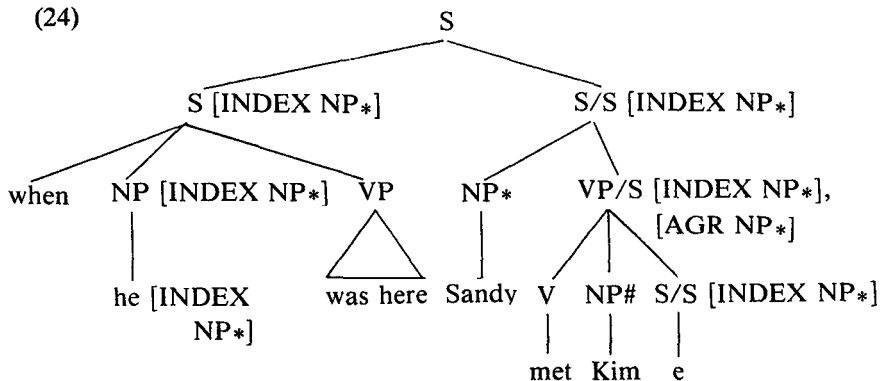
In a few paragraphs so far, we have been concerned about the question of whether a direct object NP can control a time adverbial clause in a sentence like (19). In the second reading of the sentence, the direct object *John* and the subject of the time adverbial clause *he* are coreferential. I believe that there is a sense in which an object NP indeed controls an adverbial clause immediately dominated by the VP of the main clause. If this is correct, the second reading of (19) can be described as usual, with no additional provision, in accordance with the FFP and the CAP.

Next we consider sentences like (23) in which adverbial clauses occur sentence-initially:

- (23) When *he* was here, *Sandy* met Kim.

It is well known that backward pronominalization is possible in sentences like (23), where the pronoun appears before its antecedent. (Recall that this was impossible in sentences like (8b).)

To handle these cases, it seems to me to be necessary to assume that adverbial clauses occurring before the main clause are topicalized constituents. With this assumption accepted, the CAP and the FFP work quite well as in all the previous cases. The structure of (23) may be represented as follows:



First, within the topicalized clause, the INDEX feature specification is percolated up to the top S no matter what value it may be. The Immediate Dominance Rule introducing topic constituents (i.e., $S \rightarrow XP, S/XP$) stipulates that the topicalized S must be the same as the SLASH category of S/S dominated by the top S. While this SLASH S is licensed by the ID rule, the other two SLASH Ss are instantiated by the FFP. Licensed or instantiated, SLASH categories must have identical features, so that all the three SLASH Ss have the same value for the INDEX feature, whatever the value may be. Now, crucially, the value of the INDEX feature is determined within the VP/S constituent by the operation of the CAP: the original CAP rules that the AGR value of VP/S be identical to the subject NP*, and the CAP (Addenda) ensures that the value of the INDEX feature of S/S be the same as the value of its mother's (i.e., VP/S's) AGR features. Thus the tree (24) meets the FFP and CAP as well as other stipulations imposed by ID rules.¹¹

Note that the object NP# *Kim* may control S/S. Then the INDEX feature value should be NP# rather than NP* according to the CAP. But, then, NP#, containing a female feature specification, would clash with the male feature specification. This feature conflict would make all the SLASH Ss illegal categories, which would explain the unacceptability of the sentence.

6. Summary

In this paper, I have sketched a GPSG account of the agreement phenomena between pronominal anaphors and their antecedents in English. The account has the following properties:

1. It employs the FOOT feature INDEX obeying the Foot Feature Principle and a slightly enriched version of the Control Agreement Principle.
2. It shows how the distribution of INDEX feature specifications is regulated in a tree, and that this is sufficient for stating the constraints on anaphoric agreement to be stated.

¹¹ The same account does not seem to apply to forward pronominalization as in *When 'Sandy' was here, 'he' met Kim*. Similarly, forward or backward pronominalization involving relative clauses seems to be a problem: *The woman who hired 'Sandy' will see 'him' tomorrow*. This seems to suggest that this type of pronominalization is quite a different phenomenon from what we have treated in this paper. It leads me to speculate that 'pronouns' occurring before 'antecedents' in those circumstances might not really be a pronoun, but some kind of name. Conversely, 'antecedent' names occurring before 'pronouns' in those circumstances might not really be a name in an ordinary sense, but some kind of anaphor. But this speculation should not be taken seriously at the moment. Anyhow, I will not pursue this matter here, leaving it for future study.

3. It is basically a syntactic explanation, though linked with some semantic concepts, especially those of the theory of semantic types developed in GPSG.
4. It assumes that direct (or indirect) objects control *that*-clause or *when*-clause complements.
5. It treats sentence-initial adverbial clauses as topicalized.
6. It does not appeal to variable binding or indexing devices, nor to levels of 'deep structure' or 'logical form.'

Finally, I wish to make it clear that the account presented in this paper is to be understood as exploratory approach, not anything like a definitive solution to the problem of pronominalization. As noted in footnotes, several problems are yet to be solved. I believe this account will be a base on which those problems are attacked.

REFERENCES

- Bach, E. (1970) 'Problominalization,' *Linguistic Inquiry* 1.
- Chomsky, N. (1981) *Lectures on Government and Binding*, Foris, Dordrecht.
- Gazdar, G., E. Klein, G. K. Pullum and I. A. Sag (1985) *Generalized Phrase Structure Grammar*, Basil Blackwell, Oxford.
- Gunji, T. (1984) 'A Phrase Structural Analysis of the Japanese Language,' Ms.
- Klima, E. (1964) 'Relatedness between Grammatical Systems,' *Language* 40.
- Langacker, R. (1969) 'Pronominalization and the Chain of Command,' in Reibel, D. A. and S. S. Schane, eds., *Modern Studies in English*, Prentice Hall, Englewood Cliffs.
- Park, B.-S. (1985) 'Some Control Agreement Problems in Korean: A GPSG Analysis of Honorific Expressions,' in S. Kubo, ed., *Proceedings of '84 Matsuyama Workshop on Formal Grammar*, The Logico-Linguistic Society of Japan, Tokyo University of Science, Tokyo.
- Pollard, C. J. and I. A. Sag (1984) 'Reflexives and Reciprocals in English: an Alternative to the Binding Theory,' in M. Barlow, D. Flickinger and M. Wescoat, eds., *Proceedings of the Second West Coast Conference on Formal Linguistics*, Stanford Linguistics Department.
- Pullum, G. K. (1985) 'Logic, Syntax, and Grammatical Agreement,' in S. Kubo, ed., *Proceedings of '84 Matsuyama Workshop on Formal Grammar*, The

- Logico-Linguistic Society of Japan, Tokyo University of Science, Tokyo.
- Ross, J. R. (1969) 'The Cyclic Nature of English Pronominalization,' in Reibel and Schane.
- Sag, I. A. (1983) 'On Parasitic Gaps,' *Linguistics and Philosophy* 6.
- Sells, P. (1984) *Syntax and Semantics of Resumptive Pronouns*, Ph.D. Dissertation, University of Massachusetts.
- Soames, S. and D. M. Perlmutter (1979) *Syntactic Argumentation and the Structure of English*, University of California Press, Berkeley.

Department of English
Kyung Hee University
1 Hoegi-dong, Dongdaemun-ku
Seoul 131
Korea