

ARIE VERHAGEN, FREE UNIVERSITY, AMSTERDAM

ON THE E-HYPOTHESIS OF DISLOCATION AND CONDITIONS ON DISCOURSE GRAMMAR

0. *Introductory Remarks*\*

Following and amending Williams (1977), I take "Discourse Grammar" to be concerned with rules whose relevant terms (e.g., deletion site and antecedent) are not necessarily contained within a single sentence, or more generally: within a single syntactic category<sup>1</sup> (cf. Williams 1977, 102, fn. 2). The aim of this paper is to investigate the question where sentence grammar ends and discourse grammar begins with respect to certain phenomena, the possibilities of formulating some general properties of discourse processes, and some possible internal structure of discourse grammar<sup>2</sup>.

The main topic of the paper will be the status of dislocation constructions and especially the view that dislocated phrases are generated as satellites of  $\bar{S}$  under a non-recursive initial symbol E. E, for *Expression*, was introduced by Ann Banfield in 1973, mainly in order to account for such things as exclamations, quotation, and the like. But several times since then, E has also been used to describe other root-constructions, e.g. by Emonds (1976), by Van Riemsdijk & Zwarts (1974) and Hirschbühler (1974), involving such phenomena as parentheticals and dislocation. Recently, Jan Koster (1978) has proposed E as a non-recursive alternative to  $\bar{S}$ , from Chomsky (1977)<sup>3</sup>. A considerable part of what I want to do now, consists of explicitly calculating the consequences of the E-hypothesis of dislocation, and I hope to show that E can better be taken as specifying a certain level in discourse grammar, than as an initial symbol for the phrase structure rules in sentence grammar.

1. *Right Dislocation and the E-hypothesis*

The problem leading to the considerations to be presented is exemplified by such sentences as (1).

- (1) Niemand geloofde het bericht dat Jan ziek was gisteren  
 Nobody believed the message that John ill was yesterday

{ "Yesterday, nobody believed the message that John was ill" }  
 { "Nobody believed the message that John yesterday was ill" }

This sentence is ambiguous; the adverbial *gisteren* ("yesterday") either belongs to the complete sentence, as indicated in the first gloss, or to the embedded sentence, giving an interpretation as in the second gloss. The problem concerns the structure and the derivation of (1). There are two views in the literature. One states that an adverbial to the right of the verb, as in (1), is set off from the rest of the sentence by 'comma intonation' and sees it as a dislocation construction (cf. Booij 1974, 634/5, who does not use the term "dislocation", though). The second assumes that the ambiguity of (1) is present in its syntactic structure and postulates a rule of Adverbial Extraposition, such that (1) has both two different deep and surface structures. The surface structures assumed are given in (2)a and (2)b, respectively, with *e* indicating the original position of the adverbial. The alleged rule is formulated in (3) (cf. De Haan 1976).

- (2)a [  $\bar{S}$  Niemand geloofde *e* [  $\text{VP}$  [  $\text{NP}$  het bericht dat Jan ziek was ] ] gisteren ]  
 b [  $\bar{S}$  Niemand geloofde [  $\text{NP}$  het bericht [  $\bar{S}_2$  dat Jan *e* [  $\text{VP}$  ziek was ] gisteren ] ] ]

(3) ADVERBIAL EXTRAPOSITION

X, AdvP, Y, VP, Z  
 1 2 3 4 5  $\Rightarrow$   
 1 e 3 4+2 5

I think that the dislocation analysis is correct, but unfortunately it is difficult to decide the matter by evidence concerning the immediate predictions that can be derived from the conflicting views, namely the question whether adverbials to the right of the verb position are possible in subordinate clauses not rightmost in the root (cf. Emonds 1976, 155). The problem is that judgments vary considerably, as will be clear from the examples (4) - (9).

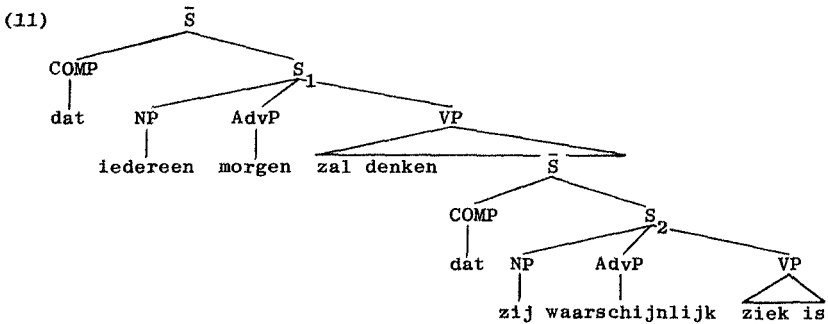
- (4)a Niemand heeft het bericht dat Jan gisteren ziek was geloofd  
 Nobody has the message that John yesterday ill was believed  
 "Nobody believed the message that John yesterday was ill"

- b \*?Niemand heeft het bericht dat Jan ziek was gisteren, geloofd  
Nobody has the message that John ill was yesterday believed
- (5)a Iedereen is door het bericht dat Jan vandaag ziek is nogal van streek  
Everyone is by the message that John today ill is quite upset  
"Everyone is quite upset by the message that John today is ill"
- b ??Iedereen is door het bericht dat Jan ziek is vandaag, nogal van streek  
Everyone is by the message that John ill is today quite upset
- (6)a Het nieuws dat Jan waarschijnlijk ziek is heeft ons erg verontrust  
The news that John probably ill is has us very disturbed  
"The news that John probably is ill disturbed us very much"
- b ??Het nieuws dat Jan ziek is waarschijnlijk, heeft ons erg verontrust  
The news that John ill is probably has us very disturbed
- (7)a Dat je misschien een beetje moe bent doet niet terzake  
That you perhaps a bit tired are does not to-matter  
"That you are perhaps a bit tired is irrelevant"
- b ?Dat je een beetje moe bent misschien, doet niet terzake  
That you a bit tired are perhaps does not to-matter
- (8)a dat als het er morgen op aankomt niemand ons enige hulp zal bieden  
that if it there tomorrow up to-comes nobody us any help will offer  
"that nobody will offer us any help, if it comes to the point tomorrow"
- b dat niemand als het er morgen op aankomt ons enige hulp zal bieden
- c dat niemand ons als het er morgen op aankomt enige hulp zal bieden
- (9)a ?dat als het er op aankomt morgen, niemand ons enige hulp zal bieden  
... if it there up to-comes tomorrow, ...
- b ??dat niemand als het er op aankomt morgen, ons enige hulp zal bieden
- c \*?dat niemand ons als het er op aankomt morgen, enige hulp zal bieden

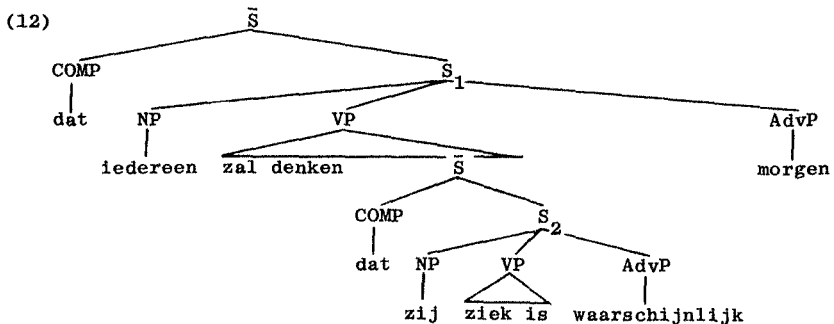
However, the generalization that is present in these data, I think, is that in all cases there is some impression of broken speech and that the most acceptable cases are those which involve a subordinate clause near the top of the dominating S or  $\bar{S}$  (cf. the worst case (4)b versus (7)b and the increasing unacceptability of the adverbial clauses in (9)). Under the dislo-

cation analysis, this is not unnatural, as it is to be expected that sentences are most easily broken up at the main constituent boundaries. But of course, one could hardly make an argument out of this. Still I think the dislocation analysis is correct, and there are other arguments for it. For example, consider the grammatical sentence (10) and especially its underlying structure (11).

- (10) dat iedereen morgen zal denken dat zij waarschijnlijk ziek is  
 that everyone tomorrow will think that she probably ill is  
 "that tomorrow everyone will think that she is probably ill"



In (11), both  $S_1$  and  $S_2$  contain a configuration which satisfies the structural condition of Adverbial Extraposition (3). So as with other Extraposition rules, the structure (12) will be derived by application of (3) in both S-domains.



However, the resulting sentence (12)a is ungrammatical.

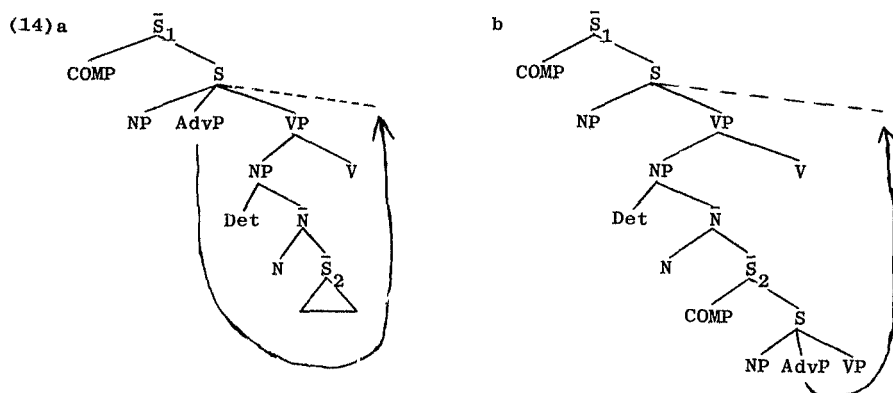
(12)a \*Iedereen zal denken dat zij ziek is waarschijnlijk morgen

It is possible to make (12)a more acceptable, namely by pausing between the two adverbials. The total sequence then has more interpretations than allowed by the structure (12); e.g., *waarschijnlijk* ("probably") may be taken to modify the entire root sentence. Both points indicate that the second adverbial is then to be regarded as a more or less independent expression<sup>4</sup>; thus, there is no grammatical sentence corresponding to (12).

In the dislocation analysis, the ungrammaticality of (12)a is no surprise, because only one S-level, namely the root, is involved. So let us suppose that the dislocation analysis is essentially correct (we will encounter more evidence later on). In a framework as Emonds (1976), where adverbial dislocation would be analysed as a movement rule, the ambiguity of (1) need not be a problem; though the sentence would have only one surface structure, something like (13), it could have two deep structures with different positions of the adverbial, accounting for the ambiguity.

(13) [ $\bar{S}_1$  Niemand geloofde [ $_{NP}$  het bericht [ $\bar{S}_2$  dat Jan ziek was ]]] gisteren ]

Thus, this 'classical' dislocation rule would operate as indicated in (14).



But this gets us into theoretical problems, as will be clear from (14)b. The adverbial would be extracted from a tensed clause, from a position in the domain of a specified subject and in a complex noun phrase, violating the Propositional Island Condition (PIC), the Specified Subject Condition (SSC)

and Subjacency. We would definitely not allow such a remarkable process, unless under very heavy empirical pressure and no alternative analyses available. The alternative is, of course, to have the dislocated adverbial generated by the base rules, together with an interpretive rule that could optionally relate the adverbial to the *entire* embedded S, so that its subject and tense cannot come into play. Subjacency is no problem either, as it is not supposed to hold for interpretive rules. Now note that the base rule involved cannot be something like (15), a rewriting rule for S.

(15)  $S \rightarrow NP VP AdvP$

This is impossible because S is recursive<sup>5</sup> and the ill-formed structure (12) would be base-generated if (15) were in the grammar. So we have to use a non-recursive initial symbol, i.e. E, and then we might get something like rule (16) where X can be spelled out as any major category and n indicates the maximal projection (number of bars) in terms of X-bar theory; (16) accounts for both left and right dislocation.

(16)  $E \rightarrow X^n \bar{S} X^n$

Now, in fact there is positive evidence that right dislocated adverbials are not contained in an S or  $\bar{S}$ ; this evidence of course also favors the dislocation analysis over the extraposition analysis. For example, they are indeed, as implied by (16), in complementary distribution with right dislocated NP's and PP's. See the examples in (17) and (18), which are parallel to (12)a.

(17)a \*De generaal zal het wel kunnen verwerken waarschijnlijk de nederlaag  
 The general will it can assimilate probably the defeat  
 "The general will probably get over it, the defeat"

b \*De generaal zal het wel kunnen verwerken de nederlaag waarschijnlijk  
 The general will it can assimilate the defeat probably

(18)a \*Marie heeft er nooit van gehouden helaas van zwarte olijven  
 Mary has there never of loved unfortunately of black olives  
 "Mary unfortunately never liked them, black olives"

b \*Marie heeft er nooit van gehouden van zwarte olijven helaas  
 Mary has there never of loved of black olives unfortunately

Again, as with (12)a, these cases can be improved by making the second phrase more independent by means of a pause. I will get back to that below. The point here is that adverbial dislocation should - if possible - be identified with right dislocation of NP and PP, because of their being in complementary distribution. Now dislocation of NP and PP cannot be a movement rule; the arguments that Van Riemsdijk & Zwarts (1974) give concerning left dislocation, can almost all be carried over to right dislocation; e.g., reflexive and reciprocal NP's may not be dislocated (while they can undergo the root transformation of Topicalization). Note also that right dislocation as a movement rule would cause violations of the principle of strict cyclicity. For instance, the rule that moves the r-pronoun (cf. Van Riemsdijk 1978) in (19) could not operate on the  $S_2$  cycle until the PP had been dislocated.

(19) [ $\bar{S}_1$  Ik denk [ $\bar{S}_2$  dat Marie er niet e van houdt ] van zwarte olijven ]

I think that Mary there not of loves of black olives

"I think that Mary doesn't like them, black olives"

As dislocation would be a root transformation, i.e. operating on the  $S_1$  cycle in (19), the r-pronoun is only present when  $S_2$  has already been passed in the cycle; therefore, right dislocation of PP cannot be a movement rule, assuming strict cyclicity.

So if adverbial dislocation is to be identified with dislocation of NP/PP, and if the latter is not a movement rule, then adverbial dislocation cannot be movement either.

Another point concerns stress. In general, adverbials may be stressed, also if they are final in a sentence; cf. (20).

(20) De minister ontvangt de delegatie morgen

The secretary receives the delegation tomorrow

"The secretary will receive the delegation tomorrow"

This sentence can be derived by verb placement (cf. note 2) from the underlying (21).

(21) (dat) de minister de delegatie morgen ontvangt

(that)the secretary the delegation tomorrow receives

A dislocated adverbial cannot have stress, however. I.e., stressed adverbials to the right of any verb in subordinate clauses and to the right of non-finite verbal elements in root sentences are impossible; all cases in (22) are out.

- (22) a \*dat de minister de delegatie ontvangt morgen  
 that the secretary the delegation receives tomorrow
- b \*dat de minister de delegatie zal ontvangen morgen  
 that the secretary the delegation will receive tomorrow
- c \*De minister [<sub>v</sub> zal ] de delegatie [<sub>v</sub> e ] ontvangen morgen  
 The secretary will the delegation receive tomorrow

Dislocated adverbials share this property with other right dislocated phrases:

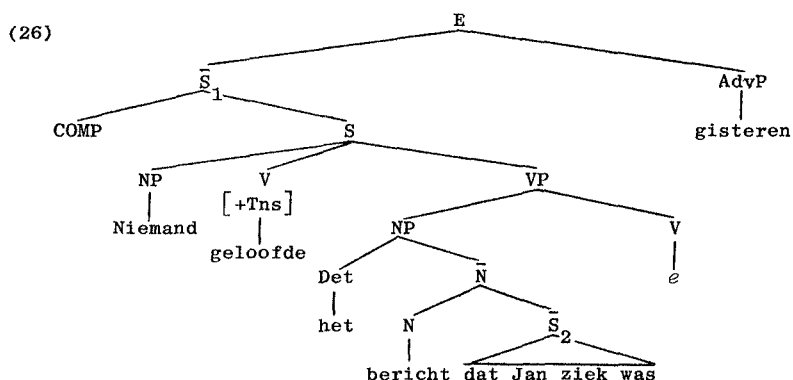
- (23) \*De generaal zal het wel kunnen verwerken de nederlaag  
 The general will it can assimilate the defeat  
 "The general will get over it, the defeat"

They do not share it with complements of S or VP to the right of the verb; these may have stress, as can be seen from (24) and (25).

- (24) dat de delegatie zal vertrekken na ontvangst door de minister  
 that the delegation will leave after reception by the secretary  
 "that the delegation will leave after being received by the secretary"
- (25) dat de minister de delegatie opwachtte op het vliegveld  
 that the secretary the delegation awaited at the airport  
 "that the secretary waited for the delegation at the airport"

Thus, the ungrammaticality of the sentences in (22) and (23) cannot be attributed to some general constraint against stress to the right of the verb; apparently, the generalization is that phrases not contained in the sentence may not bear sentence stress.

So let us assume that the E-hypothesis is correct. Then the structure of sentence (1) would be as in (26).



But now we see that we have not succeeded in solving our problems with the conditions on grammatical rules, precisely because we had to assume that the adverbial is not contained in any S, so not in  $S_1$  either. The rule associating *gisteren* ("yesterday") in (26) with the embedded sentence  $S_2$  would violate PIC and SSC because of the tense and the subject of  $S_1$ ; i.e., it would be able to relate positions as those of A and B in structures of the form (27) (where B is in the domain of the subject NP and tense of S), which should be impossible because of PIC and SSC<sup>6</sup>.

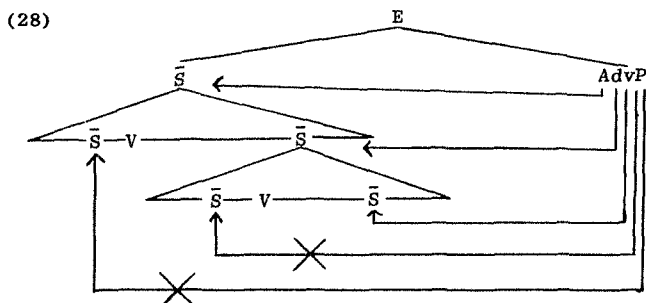
(27) ... [<sub>S</sub> ... NP ... Tense ... B ... ] ... A ...

Thus, it seems that we are forced to the conclusion that right dislocation always may violate conditions on rules of grammar. This calls into question the status of the symbol E, as it comes down to saying that rules of E-interpretation may violate the conditions. However, in (16) E is taken as an initial symbol for the phrase structure rules and in general, rules whose domains are limited by syntactic categories obey the conditions. So some doubt arises: is E a syntactic category? And if not, what else could it be, then?

For one thing, note that it is very unclear how E would fit into some not completely unrestricted theory of phrase structure, say like X-bar theory. But there is more to it.

Under the E-hypothesis, sentence (1) has only one syntactic derivation; (26) is base-generated. Therefore, the burden of accounting for the ambiguity of (1) is now completely on the interpretive rule involved. Thus, this rule

must be able to associate the adverbial with several S's contained in the root (including the root itself), but not all of them, of course. Possible and forbidden configurations are as indicated in (28).



The point is that association may only occur with S's rightmost in the root. This is a well known fact; it is implicit in such statements from Emonds as that nothing can be concluded from examples involving right dislocation from embedded clauses rightmost in the root (e.g., Emonds 1976, 33, 155). So the rule of E-interpretation for associating a dislocated adverbial with some sentence might look something like (29)<sup>7</sup>.

(29)  $X, \bar{S}, AdvP, Y \rightarrow X ((\bar{S}) AdvP) Y$

As we have seen, this rule is not subject to constraints involving the structural position (in terms of hierarchy) of the  $\bar{S}$  with respect to the AdvP; the only relevant condition is that these be adjacent, as is stated in (29) itself. Now something the same applies to right dislocation of NP's and PP's (again, not a very new observation in itself): the S's dominating the pronouns coreferential with NP or PP must be rightmost in the root; thus we have the pattern in (30).

(30)a Ik haat [<sub>NP</sub> het idee [<sub>S</sub> dat ik het op moet eten]] dat konijn  
 I hate the idea that I it up must eat that rabbit  
 "I hate the idea that I have to eat it, that rabbit"

b \*Ik zal [<sub>NP</sub> het idee [<sub>S</sub> dat ik het op moet eten]] blijven haten, dat konijn  
 I will the idea that I it up must eat stay hate that rabbit  
 "I will always hate the idea that I have to eat it, that rabbit"

c Ik zal het idee blijven haten [ $\bar{S}$  dat ik het op moet eten] dat konijn  
 I will the idea stay hate that I it up must eat that rabbit

In (30)a, the embedded clause (in a complex NP') is rightmost in the root, as a consequence of verb placement and the absence of other, non-finite verbal elements, right dislocation, i.e., coreference of *het* ("it") and *dat konijn* ("that rabbit"), is possible. In (30)b, there are infinitives to the right of the embedded clause containing the pronoun and the sentence is out. If the embedded clause is extraposed from the NP and thus gets into rightmost position, right dislocation is possible, as can be seen from (30)c. So what we have here is some general condition on the interpretation of right dislocated phrases (RDP's). As the rule for NP's will probably mention a pronoun and the NP (though this is perhaps not necessary), I assume the conventions in (31) for the interpretation of the notion "involve", to be able to state the condition uniformly.

(31)a A node  $\alpha$  is involved in a rule if it is analyzed by a constant in the structural condition of the rule,

b if  $\beta$  dominates  $\alpha$ , and  $\alpha$  is involved in a rule, then  $\beta$  is involved in that rule.

We can now state (32), where as a consequence of (31)b,  $\bar{S}$  is also taken to be involved in a rule if some node dominated by  $\bar{S}$  is involved in it, and  $e$  means empty of terminal (lexical) elements.

(32) ADJACENT S CONDITION on right dislocation (ASC)

No E-rule involves  $\bar{S}$ , X in

...  $\bar{S} \alpha X$  ...

unless  $\alpha = e$

As I said, ASC has been implicit in much work on dislocation, especially Emonds', but for some reason, it was - to my knowledge - never stated explicitly and one never really considered its consequences.

One immediate consequence is that no more than one dislocated phrase is allowed, for if there are more than one,  $\alpha$  in (32) will not be empty with respect to at least one phrase (X), as it will contain another dislocated

phrase. Thus, we do not need a rewriting rule for E to stipulate that there is only one 'dislocation position', i.e., *ASC* appears to make E superfluous, at least in this function. So E is not only suspicious as a syntactic category, it also seems to be redundant. Apparently, we used it above to express certain non-syntactic relations; in (16), however, this is done by means of a phrase structure rule, i.e., a rule of the same type as syntactic base rules; this is at best confusing, as we now see. In my opinion, the diagnosis of this confusion is perfectly worded by Chomsky in discussion with generative semantics:

Virtually anything can be expressed as a phrase marker, i.e., a properly parenthesized expression with parenthesized segments assigned to categories (Chomsky 1972, 125).

So suppose we drop E. This means that an  $\bar{S}$  and a dislocated phrase are not united in some syntactic structure; they are merely a linear sequence of two syntactic categories, a sentence<sup>8</sup> and a bare phrase, i.e., a part of a discourse. This is at least convenient in this respect, that the rule associating a phrase and a sentence would automatically be a rule of discourse grammar - recall that I defined discourse grammar rules as rules whose relevant terms are not (necessarily) contained within a single syntactic category - and as such not subject to conditions on rules of sentence grammar. Though this view will need some modification, I would first like to speculate a bit more in this direction.

## 2. Excursion: "Free" Deletion and Right Dislocation

We can say that the discourse rules involved in right dislocation, though not bound by sentence grammar conditions, nevertheless do not apply freely. They are bound by their own kind of conditions, so to speak, namely *ASC*. In turn, this means that *ASC* is a condition on certain rules of discourse grammar. Thus, the question arises as to its scope. Indeed, something like *ASC* seems to be valid for discourse rules involving deletion, like those discussed in Williams (1977), so called free deletion rules. Dutch does not have a rule of VP-deletion, but we do have sluicing. The antecedent of a sluiced constituent may be in an embedded clause, but

this clause must be rightmost in the root; i.e., (33) shows the same pattern as (30)<sup>9</sup>.

- (33)a Piet verspreidde [<sub>NP</sub> het gerucht [<sub>S</sub> dat iemand mij wilde vermoorden ]]  
 Peter spread the rumor that someone me wanted kill  
 "Peter spread the rumor that someone wanted to kill me"  
 Maar hij wist niet wie --- .  
 But he knew not who  
 "But he didn't know who"
- b ??Piet heeft het gerucht [<sub>S</sub> dat iemand mij wilde vermoorden ] verspreid  
 Peter has the rumor that someone me wanted kill spread  
 /Maar hij wist niet wie --- .  
 But he knew not who
- c Piet heeft het gerucht verspreid [<sub>S</sub> dat iemand mij wilde vermoorden ]  
 Peter has the rumor spread that someone me wanted kill  
 Maar hij wist niet wie --- .  
 But he knew not who

In (33)b, the embedded clause is not rightmost in the root, and apparently Sluicing cannot find an antecedent in it; (33)c, with extraposition of  $\bar{S}$  from NP, is far better than (33)b, precisely as with (30)c versus (30)b. Another instance is given in (34).

- (34)a ??Piet vertelde het verhaal [<sub>S</sub> dat iemand mij wilde vermoorden ] aan Jan  
 Peter told the story that someone me wanted kill to John  
 "Peter told the story that someone wanted to kill me to John"  
 /Maar hij vertelde er niet bij wie --- .  
 But he told there not with who  
 "But he didn't tell who"
- b Piet vertelde Jan het verhaal [<sub>S</sub> dat iemand mij wilde vermoorden ]  
 Peter told John the story that someone me wanted kill  
 "Peter told John the story that someone wanted to kill me"  
 Maar hij vertelde er niet bij wie --- .  
 But he told there not with who  
 "But he didn't tell who"

In (34)a, it is not a participle (as in (33)), but a prepositional indirect object that intervenes between the clause containing the antecedent and the one containing the deletion site; the discourse is ill-formed, i.e., observationally, the second sentence sounds 'unfinished'. But in (34)b, the subordinate clause is final in the root, because the indirect object is not prepositional and thus precedes the direct object<sup>10</sup>, and (34)b is good. Now it cannot be *ASC* itself that is involved in (33) and (34), because a dislocated phrase may occur between the two sentences; so  $\alpha$  of (32) would not be empty; see (35), for example.

- (35) Piet heeft 't gerucht verspreid [<sub>S</sub>-dat iemand me wilde vermoorden] gisteren  
 Peter has the rumor spread that someone me wanted kill yesterday  
 Maar hij wist niet wie --- .  
 But he knew not who

But here we have, I think, a key to the real nature of E. Though this nature is not syntactic, it is still a fact that a dislocated phrase is felt to form some unit with the preceding  $\bar{S}$ ; furthermore, a dislocation construction is speaker-bound (contrary to discourse rules involving deletion) and dislocated phrases are not independent units of intonation, but they are pronounced with a continuation of the low pitch that occurs after the final stress in the sentence<sup>11</sup>. We might say then, that as a result of an interpretive rule of dislocation (cf. (29)), an interpretive unit, i.e., an "expression", is *created*. I.e., E would not exist in the grammar, but it would at some level of interpretation. Then it may be present for the discourse rules involving deletion (for shortness: D-rules). This amounts to distinguishing at least two levels within discourse grammar, one of E-rules and one of D-rules.

We can now represent the situation of (35) as in (36).

- (36)  $\left| \begin{array}{c} E_1 \\ \left[ \bar{S} \dots \left[ \bar{S} \dots \right] \right] \text{ AdvP} \end{array} \right| \left| \begin{array}{c} E_2 \\ \left[ \bar{S} \dots \left[ \bar{S} \dots \right] \right] \end{array} \right|$

The sequences  $\bar{S} \text{ AdvP}$  and  $\bar{S}$  are interpreted as expressions and at that level of interpretation these are adjacent. So we might formulate the conditions in (37) and (38).

(37) ADJACENT E CONDITION (*AEC*)

No D-rule involves  $E_1, E_2$  in

$$\dots E_1 \alpha E_2 \dots$$

unless  $\alpha = e$

(38) ROOT S BOUNDARY CONDITION (*RSBC*, first version)

No D-rule involves  $\bar{S}, X$  in

$$\dots \bar{S} \alpha ]_{\bar{S}} \dots X \dots$$

unless  $\alpha = e$

The effects of the Root S Boundary Condition<sup>12</sup> are illustrated by (33) and (34). The Adjacent E Condition is intended to allow for (35), but in a restricted way, the restriction being that no complete expression may occur between two positions involved in a deletion process of discourse grammar. This seems to be correct, given the difference between (39) and (40).

- (39) A: Ik hoor net [ $\bar{S}$  dat de politie iets ontdekt heeft] gisteren  
 I hear just that the police something discovered has yesterday  
 "I just heard that the police discovered something yesterday"
- B: Weet je ook wat --- ?  
 Know you also what  
 "Do you also know what?"

- (40)?? A: Ik hoor net [ $\bar{S}$  dat de politie iets ontdekt heeft ] gisteren  
 B: Het zou tijd worden. ~~Weet je ook wat --- ?~~  
 "It was about time. Do you also know what?"

(39) is the same as (35), i.e., it shows the pattern (36), which is allowed by *AEC* and *RSBC*. But in (40), there is a complete expression between the terms involved in the deletion; *RSBC* is satisfied, but *AEC* is not, so Sluicing cannot find an appropriate antecedent.

For the present discussion of dislocation, however, something else is more interesting. Recall that I said above, in connection with (12)a and (17), (18), that one could make a sequence of two dislocated phrases more acceptable by pausing. It seems rather natural to suppose that the dislocated

phrase then counts as an independent E in the sense we have by now established. Given AEC, it follows that such an independent 'dislocated' phrase must block the application of D-rules. In fact, it does; thus, (41) contrasts with (39), although the utterance of A in (41) is in itself acceptable.

- (41)?? A: Ik hoor net [<sub>S</sub>-dat ze iets ontdekt hebben] de politie - gelukkig  
 I hear just that they something discovered have the police-fortunately  
 "I just heard they discovered something, the police - fortunately"
- ≠ B: Weet je ook wat --- ?  
 "Do you also know what?"
- 

Another prediction is, that sequences of more than two bare phrases following a sentence will be absolutely impossible, no matter how they are pronounced. This is borne out, too; cf. (42).

- (42) \*Ze hebben iets ontdekt, de politie - gelukkig - gisteren  
 They have something discovered, the police - fortunately - yesterday

Note that there is no need to assume a recursive category E, or  $\bar{S}$ , to account for the possibility - under certain conditions - of two RDP's with one sentence. Moreover, the above analysis would be superior to such a proposal, as it predicts, without further stipulations, the sharp boundary between two as opposed to three dislocated phrases with regard to acceptability.

### 3. Left Dislocation

The preceding discussion has been based solely on right dislocation, but in fact the same kind of phenomena can be found at the left hand side of sentences. Most importantly, Dutch has two types of left dislocation. Consider (43) and (44).

- (43)a Dat gezeur dat kan ik niet uitstaan  
 That puling that can I not endure  
 "That puling I can't bear"
- b Dat gezeur - Ik kan het niet uitstaan  
 That puling - I can it not endure  
 "That puling - I can't bear it"

- (44)a Van Agt die            krijgen we er        wel        onder  
 Van Agt that-one get        we there indeed under  
 "Van Agt we will overcome"
- b Van Agt - we krijgen hem er        wel        onder  
 Van Agt - we get            him there indeed under  
 "Van Agt - we will overcome him"

The a-cases are to be pronounced without a pause (though they are not ungrammatical with one); there is a singular intonational contour which includes the dislocated NP. The b-cases have a pause between the bare phrase and the sentence, and these have distinct, independent intonational contours; if the b-cases are pronounced without a pause, with a single intonational contour, they are ungrammatical.

These differences have been noted before; in particular, I refer to a structuralist grammarian of Dutch, Paardekooper (1977). He calls the b-cases ungrammatical, i.e., excluded from the class of sentences, where his concept of a sentence (especially, "sentence in its broader sense") comes close to our "expression", i.e. the interpretive E. Furthermore, he notes that the a-cases contain an anaphoric pronoun, controlled by the left dislocated phrase (short: LDP) and that this pronoun must be leftmost in the sentence, i.e., adjacent to the LDP (Paardekooper 1977, 46/47). Paardekooper describes these pronouns - correctly, I think - as 'indicating linguistic control' in these positions, in contrast to other cases and other pronouns, which indicate "situational" (pragmatic) control. There is a separate class of words in Dutch that can have the function of indicating linguistic control by a phrase not contained in the sentence; all but one of them begin with a *d*; so for ease of reference we may call them the class of d-words<sup>13</sup>. Apparently, then, in the case of left dislocation it is the anaphoric d-word itself that must be adjacent to the LDP in order for the sequence of a bare phrase and a sentence to be interpretable as an expression. The b-cases of (43) and (44) are then considered to be sequences of two E's. To express this, and especially Paardekooper's insights, we can formulate the following condition<sup>14</sup>:

(45) ADJACENT ANAPHOR CONDITION (*AAC*)

No E-rule involves X, Y in

$$\dots X \alpha \left[ \bar{S} \beta Y \dots \right]$$

unless  $\alpha = \beta = e$

Left dislocation also shows the properties of right dislocation that followed from the *AEC*, namely, first, that there may be two LDP's, one of them being an independent E, with its own intonational contour and separated from the other one by a pause, and second, that there can never be more than two LDP's. These properties are illustrated by (46), (47) and (48).

(46) Van Agt - Den Uyl die kan hem niet uitstaan  
 Van Agt - Den Uyl that-one can him not endure  
 "Van Agt - Den Uyl can't stand him"

(47) \*Den Uyl - Van Agt - de een kan de ander niet uitstaan  
 Den Uyl - Van Agt -the one can the other not endure

(48) \*Den Uyl - Van Agt - dat gezeur dat kunnen ze niet uitstaan  
 Den Uyl - Van Agt - that puling that can they not endure

The main difference between *AAC* and *ASC* is that *ASC* only requires that the dominating  $\bar{S}$  is adjacent to the dislocated phrase. Thus, *ASC* is less restrictive than *AAC*. This might account for the differences of acceptability between left and right dislocation involving embedded clauses. In both cases, the dominating clause must be broken up, which may lead to bigger or lesser problems of understanding. But note that for right dislocation, the condition involved, *ASC*, can still be satisfied rather easily. This might be the explanation for the relative acceptability of cases like (7)b, repeated here for convenience, with a structure as indicated.

(7)b ? $\left[ \bar{S} \text{ Dat je een beetje moe bent} \right] \left[ \text{AdvP}_{\text{misschien}} \right]$ , doet niet terzake  
 That you a bit tired are perhaps does not to-matter

In the case of left dislocation, however, the LDP must not only be adjacent to the  $\bar{S}$  involved, but also to the anaphor controlled by it; but we cannot have preposing of d-words in subordinate clauses, and the COMP position must

always be filled with a lexical complementizer; thus, the cases in (49) are all completely unacceptable.

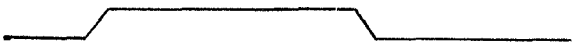
- (49)a \*Ik geloof niet dat Piet die ziek is  
 I believe not that Peter that-one ill is  
 "I don't believe that Peter is ill"
- b \*Ik geloof niet dat Piet die is ziek
- c \*Ik geloof niet Piet dat die ziek is
- d \*Ik geloof niet Piet die ziek is
- e \*Ik geloof niet Piet die is ziek

However, though everyone will judge them as not Dutch, one often hears constructions as (50) in colloquial speech.

- (50) (\*)Piet gelooft dat die plannen [ $\bar{S}$  dat we die allang kenden]  
 Peter believes that those plans that we those already long knew  
 "Peter believes that we knew those plans already for a long time"

Note that the 'LDP', though not adjacent to its d-word, is adjacent to the  $\bar{S}$  involved, so that we have, so to speak, a less serious violation of AAC than in (49): at least  $\alpha$  of (45) is empty in (50)<sup>15</sup>.

There still seems to be an unexplained asymmetry between left and right dislocation. It involves stress. The 'normal' intonational contour of a sentence with dislocated phrases is as in (51) (the contour is a so called hat-pattern).

- (51)   
 [ $X^n$  .....] [ $\bar{S}$  .....] [ $X^n$  ...]  
 Dat gezeur dat kan ik niet uitstaan vandaag  
 That puling that can I not endure today  
 "That puling I can't bear today"

I observed earlier that RDP's cannot bear sentence stress, i.e., the final change of pitch in the intonational contour (cf. Verhagen 1979). I used this as an argument against the extraposition analysis of (1). But (51) shows that an LDP can easily contain a change of pitch: an RDP may not contain final stress, but an LDP may contain initial stress. This asymmetry

is only a seeming problem, however. Dislocated phrases, both left and right, define 'domains of interpretation', i.e., frames in which the rest of the assertion, in particular the content of the  $\bar{S}$ , is to be interpreted, in other words: *topics*. It seems natural to assume that this is a function of their being outside  $\bar{S}$ ; we may consider this as the germ of truth in Chomsky's (1977) rule  $\bar{S} \rightarrow \text{TOP } \bar{S}$ .

Then, given that sentence stress (i.e., final stress) marks the focus of the sentence, it follows that a dislocated phrase cannot bear final stress; but any stress on an RDP would of course have to be final in the intonational contour; so RDP's cannot be stressed at all. Stress in an LDP, on the other hand, will be an initial stress and it has been shown on completely independent grounds, e.g., in Salverda (1979) and Verhagen (1979), that initial stress marks topic-function in the information structure of expressions. So (51) is precisely what is to be expected, in fact.

It is already well known that intonation determines such matters of interpretation as focus and topic; in the above discussion, I have argued that E is not a structural category, but an interpretive one. We see here, that these two approaches converge: the unit of intonation is not the same as the syntactic unit S ( $\bar{S}$ ), but as the interpretive unit E.

#### 4. Concluding Remarks

As a consequence of *ASC* and *AAC*, the left and right hand boundaries of a root sentence function as a kind of 'gates' through which bare phrases can be connected with elements in sentences. We have already seen that the right hand boundary of a root sentence functions in the same way with respect to D-rules, as was laid down by *RSBC* in (38). Consider now (52).

- (52)a Het was ons bekend [ $\bar{S}$  dat er iemand in het gebouw zou kunnen zijn]
- It was us known that there someone in the building would can be
- "It was known to us that someone might be in the building"
- We wisten alleen niet wie --- .
- We knew only not who
- "We only didn't know who"

b [ $\bar{S}$  Dat er iemand in het gebouw zou kunnen zijn] was ons bekend  
 That there someone in the building would can be was us known  
 "That someone might be in the building, was known to us"  
 We wisten alleen niet wie --- .  
 We knew only not who  
 "We only didn't know who"

As *RSBC* now stands, (52)b seems to be a counter-example; the subordinate  $\bar{S}$  is not rightmost in the root and yet Sluicing is not blocked. However, the subordinate  $\bar{S}$  is *leftmost* in its root in (52)b; so let us have a look at a case where the subordinate clause is not leftmost and not rightmost, e.g., as in (53)b.

(53)a We kenden de mogelijkheid [ $\bar{S}$  dat er iemand in het gebouw was ]  
 We knew the possibility that there someone in the building was  
 "We knew the possibility that there was someone in the building"  
 We wisten alleen niet wie --- .

b ??De mogelijkheid [ $\bar{S}$  dat er iemand in het gebouw was ] kenden we  
 The possibility that there someone in the building was knew we  
 "The possibility that there was someone in the building, we knew"  
~~We wisten alleen niet wie --- .~~

So suppose we change *RSBC* into (54), to account for (52) and (53), too.

(54) ROOT S BOUNDARY CONDITION (*RSBC*, tentative final version)

No D-rule involves  $\bar{S}$ , X in

... [ $\bar{S}$   $\alpha$   $\bar{S}$   $\beta$ ] ... X ...  
r

unless  $\alpha = e$

or  $\beta = e$

One might suggest that there could have been a vacuous application of extraposition from NP in (53)a and that the Complex NP Constraint could somehow be responsible for the blocking of Sluicing in (53)b, where extraposition certainly has not applied. However, the difference between (55)a and (55)b, which is the same as in (53), suggests that (54) embodies the correct

generalization.

(55)a Het gaat me te ver te zeggen [ $\bar{S}$  dat de politie iets ontdekt heeft ]  
 It goes me too far to say that the police something discovered has  
 "It's going too far for me to say that the police discovered something"  
 Want ik zou niet weten wat --- .  
 For I would not know what  
 "For I have no idea what"

b ??Te zeggen [ $\bar{S}$  dat de politie iets ontdekt heeft ] gaat me te ver  
 To say that the police something discovered has goes me too far  
 "To say the police discovered something, is going too far for me"  
~~Want ik zou niet weten wat --- .~~  
 For I would not know what  
 "For I have no idea what"

In (55)b, the complete infinitival construction is in front position, so that the  $\bar{S}$  involved is neither leftmost nor rightmost in its root, and indeed, Sluicing is blocked in just the same way as in (53), indicating that (54) is to be preferred over the other solution just suggested.

I will conclude with some general remarks about the conditions proposed in this paper. It will be clear that there is still much to be done in this respect, so it is quite likely that several parts of this analysis will have to be amended. The biggest problems, I think, are with the Root  $\bar{S}$  Boundary Condition; recall the problems of observation hinted at in note 9. Still I think that the general strategy behind this all is promising: we have to try to get over the situation that discourse grammar is only negatively defined (essentially, as not sentence grammar), and to reach a more positive formulation of properties that discourse processes have of their own. One direction for future research in these matters is suggested by the following considerations.

Note that there is a considerable overlap between the conditions, e.g., between *ASC* and *RSBC*; there also seem to be reasons to assume that *ASC* and *AAC* partly cover the same phenomena (cf. note 15). Such redundancies should be removed and in order to do that, we might look for features that are

common to the conditions. Now, these may be said to define a kind of *locality* in discourse, as they limit the space in which items involved in a process may occur. Thus, they are parallel to conditions on rules of sentence grammar (cf. Koster 1978). The crucial notions for the latter conditions are relations defined in terms of dominance, such as c-command, superiority and the like, i.e., they involve hierarchy in phrase markers. Our conditions, on the other hand, crucially involve *linear* relations, especially contiguity. It might be suggested, then, that this embodies a characteristic difference between sentence and discourse grammar, the first being essentially the theory of linguistic structure, the second of sequences of sentences and phrases.

Other points that might be noted in this connection, concern the facts that the only syntactic category explicitly mentioned in the conditions is  $\bar{S}$ , that "contained in a root  $\bar{S}$ " (perhaps "in an  $\bar{S}$ ", cf. note 12) is the only not strictly linear relation between elements that seems necessary, and that there is no nesting (which leads to the 'gate-character' of the root S boundaries, as noted). I do not intend these sketchy remarks to be taken as containing established results; I just hope that they give some indication of the way that work along these lines might progress.

March 1979

*Subdepartment of Dutch  
Free University  
1007 MC Amsterdam  
The Netherlands*

#### Footnotes

\* Research for this paper was undertaken as a part of the project "Grammar and Pragmatics", Free University Grant 76/4. I am especially indebted to Thijs Pollmann, for suggesting possible general consequences of some earlier work I did on right dislocation. I think that his stimulating remarks have strongly influenced the present form and content of the paper.

<sup>1</sup> The reason for this last amendment is that "sentence" may be used in two ways: 1st, referring to a formally defined syntactic category (what is dominated by the designated symbol S), or 2nd, as a more intuitive concept, at

the observational level. These uses are clearly related, but it is not a priori clear how; one of the implications of the work to be reported is a sophistication of the notion "sentence"; cf. note 8.

2 As this investigation proceeds on the basis of Dutch, some remarks on Dutch grammar may be in order. Dutch is underlyingly an SOV (i.e., verb final) language (cf. Koster 1975). This is the constituent order in subordinate clauses; therefore I sometimes give examples in terms of subordinate clauses. In the surface structure of root sentences, the finite verb and only the finite verb appears in second position, no matter what is in first position, an adverb, the subject, a topicalized object, or whatever other phrase. A generally recognized consequence is that most phrases are generated in the base (to the left of the verb. Only sentences, and perhaps PP's, may occur as base generated complements to the right of V (cf. Koster 1978, 11/12).

3 The argument below, concerning the status of E, thus carries over to  $\bar{S}$ , apart from recursiveness; cf. the end of section 2.

4 Thus, if we pronounce the two adverbials as parts of a single intonational pattern, (12)a is ungrammatical, and this is not due to a general constraint against adjacent adverbials, which simply does not exist in Dutch, witness the grammaticality of (i) and (ii) (for English, cf. Jackendoff 1977, 73).

(1) Jan gaat morgen waarschijnlijk naar huis  
 John goes tomorrow probably to home  
 "Tomorrow, John probably goes home"

(ii) Jan gaat waarschijnlijk morgen naar huis  
 John goes probably tomorrow to home  
 "John probably goes home tomorrow"

5 For the same reason, it cannot be a rule involving  $\bar{S}$  or  $\bar{\bar{S}}$  as a left hand symbol.

6 Note that the same problem in fact also arises if right dislocation constructions would have a surface structure as in (13), because the first embedded clause might contain another one, from which an adverbial may be dislocated; cf. (i).

- (i) [ $\bar{S}_1$  Piet gelooft [ $\bar{S}_2$  dat iedereen denkt [ $\bar{S}_3$  dat Jan ziek was] $_{3-2}$  gisteren] $_1$   
 Peter believes that everyone thinks that John ill was yesterday  
 "Peter believes that everyone thinks that John was ill yesterday"

Relating *gisteren* ("yesterday") to  $S_3$  (by movement or by interpretation) violates PIC and SSC because this relation 'crosses'  $S_2$ ; but (i) is grammatical in the relevant reading.

7 This is only a very tentative formulation. The right hand side is mainly added for concreteness; it is meant to indicate that the adverbial must be interpreted as a sentence modifier; i.e., although predicate modifiers may also be dislocated, they must then be interpreted as "domain adverbials" in the sense of Bellert (1977); see Verhagen (1979) for discussion of this use of predicate modifiers.

8 Note that this analysis leads to a 'sophistication' of the notion of a sentence, in the sense that what is a sentence in the theory, is no longer directly linked to the more intuitive notion of sentence. This is immediately clear if one considers the consequence that a grammar that does not contain E as an initial symbol for the PS rules and that contains no dislocation transformation, will simply not generate sequences as (1); according to such a theory, (1) is simply ungrammatical, i.e., *as a sentence* in the sense of the theory. This means that we will have to look for an explanation for the intuitive appearance of (1) as a 'sentence'; I will get back to that in the following sections. Such sophistication of originally vague and intuitive concepts is quite common in the development of scientific theories.

9 I mark the problematic cases with question marks, not with stars, because there are several problems in getting judgments on sequences of sentences which together exceed a certain (limited) length and degree of complexity. This manifested itself in the fact that informants became uncertain of their judgments after the presentation of the first two examples. So I presented every pair of examples involving sluicing to different informants; but this raises other questions about the reliability of the totality of the judgments. Notwithstanding these serious problems of observation, I present these data as first attempts to attain some clarity in these matters, hoping that further research may shed more real light on it.

10 It is immaterial to this discussion whether (34)b is base-generated or derived by Dative Movement; the former option is probably correct.

I have been informed that the English translations of (34) show the same pattern of acceptability, the first being bad.

11 I.e., there is normally no real pause between a sentence and a right dislocated phrase; the impression of 'comma intonation' is caused by intonational features, not by a pause in the stream of sound.

12 *RSBC* might be formulated more generally, with replacement of  $\bar{S}_r$  ("root  $\bar{S}$ ") by, simply,  $\bar{S}$ . It is clear that, at least for deletion 'over utterance boundaries', the condition will then also only be satisfied for an embedded clause if there is no dominating  $\bar{S}$  containing lexical elements to the right of the embedded  $\bar{S}$ , i.e., if the latter is not rightmost in the root. Note that *RSBC*, in either formulation, also applies to the situation where " $\alpha \bar{S} \beta$ " is identical to the  $\bar{S}$  referred to in the labeled bracket, because dominance is a reflexive relation.

13 The one exception is *toen* ("then"), which refers to past events; future events are referred to by *dan*. The class of d-words is related to the class of relative pronouns, but certainly not identical with it: e.g., there are just as many d-words that cannot be relative pronouns, as there are that can; a left dislocated clause is referred to by the d-word *dat*, but the corresponding relative pronoun is the wh-word *wat*. Therefore, I think it is not very illuminating to try - as proposed in Koster (1978) and Den Besten (1977) - to reduce Topicalization in Dutch to left dislocation plus wh- or relative pronoun movement, followed by optional deletion in COMP (only in root sentences), even apart from the considerations presented in the text to the effect that dislocation phenomena are excluded from sentence grammar. See also Van Riemsdijk (1978), 199-201. As another point, recall that reflexives and reciprocals may not be left dislocated, while they may be topicalized. The correct observation is, in my opinion, that the class of d-words indicating linguistic control (functioning as anaphors in the sense of the AAC), is identical to the class of unmarked demonstrative pro-forms (cf. Lyons 1977, 647), namely *die, dat* ("that one"), *daar* ("there"), *dan, toen* ("then"), the marked cases being *deze, dit* ("this one"), *hier* ("here") and *nu* ("now"). Thus, we have the pattern (i), \*(ii), even though the left dislocated phrase

itself contains the form *dit* ("this").

(i) Dit huis dat bevalt me wel  
This house that-one pleases me  
"This house (,it) pleases me"

(ii) \*Dit huis dit bevalt me wel  
This house this-one pleases me

Moreover, this difference indicates that one cannot use some morpho-syntactic feature [+D] to account for these phenomena, if this is to have some content in terms of morphology, i.e., as an independently identifiable feature.

- 14 Note that AAC is never satisfied unless it is satisfied for  $\bar{S} = \bar{S}_r$ ; cf. note 12.
- 15 This suggests that AAC might have to be split up into two parts, one of which is to be united with ASC.

#### Bibliography

- Banfield, A. (1973), 'Narrative style and the grammar of direct and indirect speech'. In: *Foundations of Language* 10, 1-39.
- Bellert, I. (1977), 'On Semantic and Distributional Properties of Sentential Adverbs'. In: *Linguistic Inquiry* 8, 337-351.
- Besten, H. den (1977), 'On the Interaction of Root Transformations and Lexical Deletive Rules'. Unpublished paper, University of Amsterdam.
- Booij, G.E. (1974), 'Zinsbepalingen in het Nederlands'. In: *Spektator* 3, 619-646.
- Chomsky, N. (1972), 'Some Empirical Issues in the Theory of Transformational Grammar'. In: id., *Studies on Semantics in Generative Grammar*. The Hague.
- Chomsky, N. (1977), 'On Wh-movement'. In: A. Akmajian, P. Culicover & T. Wasow (eds.), *Formal Syntax*. New York.
- Emonds, J.E. (1976), *A Transformational Approach to English Syntax*. New York.
- Haan, G.J. de (1976), 'Regelordening en domeinformuleringen op transformaties'. In: G. Koefoed & A. Evers (eds.), *Lijnen van taaltheoretisch onderzoek*. Groningen.

- Hirschbühler, P. (1974), 'La dislocation à gauche comme construction basique en français'. In: C. Rohrer & N. Ruwet (eds.), *Actes du colloque franco-allemand de grammaire transformationnelle*. Tübingen.
- Jackendoff, R.S. (1977), *X̄-Syntax: A Study of Phrase Structure*. Cambridge, Mass.
- Koster, J. (1975), 'Dutch as an SOV language'. In: *Linguistic Analysis* 1, 111-136.
- Koster, J. (1978), *Locality Principles in Syntax*. Dordrecht.
- Lyons, J. (1977), *Semantics* (Vol. 2). Cambridge.
- Paardekooper, P.C. (1977), *Beknopte ABN-syntaksis* (fifth printing). Eindhoven.
- Riemsdijk, H.C. van (1978)a, 'On the Diagnosis of Wh Movement'. In: S.J. Keyser (ed.), *Recent Transformational Studies in European Languages*. Cambridge, Mass.
- Riemsdijk, H.C. van (1978)b, *A Case Study in Syntactic Markedness: the binding nature of prepositional phrases*. Dordrecht.
- Riemsdijk, H.C. van & F. Zwarts (1974), 'Left dislocation in Dutch and the status of copying rules'. Unpublished paper, MIT/University of Amsterdam.
- Salverda, R. (1979), 'Topic, focus and preposing in Dutch declarative main clauses'. In: M. v.d. Velde & W. Vandeweghe (eds.), *Sprachstruktur, Individuum und Gesellschaft*. Tübingen.
- Verhagen, A. (1979), 'Fokusbepalingen en grammatikale theorie'. In: *Spektator* 8.
- Williams, E.S. (1977), 'Discourse and Logical Form'. In: *Linguistic Inquiry* 8, 101-139.