Tanya Reinhart
OTS Working Papers in Linguistics, 1995.

INTERFACE STRATEGIES

Contents:

0. Introduction
I. Quantifier Scope. Appeared in Linguistics and Philosophy, not included.
II. Interface economy.
III. Focus -The PF interface.
IV. Topics and the conceptual interface.

==========

0.INTRODUCTION

1. The question I want to examine here is the division of labor between the different components of linguistic knowledge.

Attempting to expand the empirical basis of syntactic theory, a substantial theoretical machinery got accumulated. The move in current syntactic theory (Chomsky's minimalist program) has been to check how much of this machinery is actually necessary. The theoretical goal is that syntactic operations - the computational system - should be driven only by purely formal and mechanical considerations, like checking morphological features. In a way, this stage is the sharpest statement of the thesis of the autonomy of syntax. We know, by now, that it is strictly impossible to derive the properties of the computational system from any functional considerations of language use. Systems of use and communication are consistent with many possible languages, and they cannot explain why the particular human language got selected. On the other hand, it is a crucial fact about human language that it can be used to argue, communicate, think, etc. If our formal analysis of the computational system turns out inconsistent with basic facts of language use, e.g. if it can be shown that the structures we generate are unusable for inference or logical entailment, this cannot be the correct analysis, since the actual sentences of human language can be used for such purpose. Capturing correctly the interface between the formal system and the systems of use is, therefore, a crucial adequacy criterion of any syntactic theory.

There is, however, no pretheoretic way to know how, precisely, the correct options of use are guaranteed in any given case, namely, how structure and use are related. Suppose we

observed, empirically, that a certain structure S is associated with a set U of possible uses. This could, in principle be explained in three ways: a. The properties necessary for U are directly encoded in S, through the computational system, as syntactic features, as specific structural configurations, or as specific conditions on derivations. b.There is no direct relation between the syntactic properties of S and U. Rather, the set U is determined solely by the systems of use. c. There are some interface strategies associating S and U, using independent properties of the CS, and of the systems of use.

Most likely, all three options exist, in fact, governing different aspects of the relations of structure and use. But the one actually favored in syntactic practice is the first - that of syntactic encoding. Many of the properties now encoded in the syntax got there in order to guarantee the correct interface with the systems of use. R(eferential), Q(uantified), F(ocus), are just a few examples. It is easy to understand this preference. Work on area c, of interface strategies (let alone on area b - the pure domain of use) is bound to be less explicit and formal than work on the computational system can be. Although lists of features (like any lists) may not be an optimal theoretical choice, they are still more explicit and precise than the vacuous narratives that one sometimes finds in discourse theory. Nevertheless, if the properties we encode in the CS do not, in fact, belong there, we are bound not to get too far. Encoding interface properties has led to an enormous enrichment of the machinery. In many cases, the result is a highly baroque syntax, which, nevertheless, fares rather poorly in capturing the interface.

Keeping in mind that we cannot know in advance what belongs where, I will focus here on the division of labor between the first and the third options above: which properties necessary for language use are directly encoded in the CS and which are governed by interface strategies.

2. It is not entirely an accident that many of the problems I will discuss evolve around questions of the properties of NPs, and their typology - particularly, the interface properties of indefinite NPs. This is, perhaps, the clearest illustration in linguistic theory of the fact that expressions do not come with their theoretical label on, and how we categorize them, is a theoretical decision. Two radically different positions exist regarding which grouping of the NPs in (1) is linguistically relevant.

1)

a. a philosopher	b. every philosopher	c. the philosopher
13 book reviews	most book reviews	Max/ Max' review

The semantic framework of generalized quantifiers, following Montague, views (1b,c) as one group - strong quantifiers - distinct from the group (1a) of weak quantifiers. There is no semantic or syntactic category, in this framework that groups (1a,b) as distinguished from (1c). In the syntactic framework, by contrast, the central division was perceived to be that between (1c) - of referential NPs, and (1a,b), of quantified NPs. The question is, then, which of these

distinctions is encoded in the computational system (as some structural property shared by members of the given NP type, or some syntactic features associated with them, etc.)

There is no doubt that the weak- strong distinction, under its semantic formulation, that will be discussed in part IV, is one of the most fruitful distinctions semantic theory has discovered. There are many known linguistic contexts which distinguish directly these two types: There sentences, extraposition, free relatives in languages like Lakhota, and many other. Furthermore, one of the most important insights of DRT is that while strong NPs are necessarily quantified, namely, their N-variable is closed internally to the NP, weak NPs may be locally open, and, thus allow for 'unselective binding'.

On the other hand, it is clear that we cannot even imagine a theory of the interface, or of language use, without entering the question of reference. The NPs of type (1c) have, obviously, different discourse uses than the quantified NPs of (1a,b), which can be witnessed also within the sentence, e.g. in the case of anaphora. Suppose we decide reference-properties need to be encoded, to guarantee the successful interface. Since the weak-strong distinction is also needed, as we saw, we end up with a three way distinction, correlating to the three groups of (1), i.e. we enriched the machinery.

Once it is assumed that referentiality is captured by syntactic encoding, the next question is what we do when we discover that, in fact, indefinites of type (1a) differ in this respect from other quantified NPs, and they allow also for what seems to be a referential use: Either introducing discourse entities, or referring back to N-sets previously mentioned in the discourse (d-linking), etc. A popular answer has been to encode this distinction as well, and assume that there are two types of indefinites, which are syntactically (and semantically) distinct. This ambiguity, again, entails a substantial enrichment of the machinery. (I will illustrate this in some detail with the proposal of de Hoop (1992) and of Diesing (1993).)

The conclusion which emerges from several of the problems in the following chapters is that the distinction which is encoded in the computational system is only the semantic distinction between weak and strong NPs ((1a) and (1b.c)).[1] The properties of indefinites which are associated with referentiality follow from (two different) interface strategies.

[1]I argued elsewhere that referential properties of NPs need not be syntactic encoded. As far as the CS is concerned, the distinction between the group (1c), of referential NPs and (1a,b) of quantified NPs, does not play any role in any of the areas for which it was argued to be needed: Neither the anaphora conditions of the CS, nor QR are sensitive to this distinction. I will not discuss these issues here.

II. INTERFACE ECONOMY.

1. Qr as a marked operation.

We saw in part I that QR is, in fact, a much more restricted operation than standardly assumed. The clearest cases of what appears as scope outside of the c-command domain, are captured, independently of QR, by the choice-function mechanism, which interprets them in situ. Still, there are cases of genuine non overt quantifier scope, for which we still need QR.

Though this QR residue can be viewed just as a standard instance of a movement operation, it still poses conceptual problems. While, as I mentioned, in part I, certain problems were always there, they are more acutely noticeable in the framework of the minimalist program. The theoretical goal is to allow movement (overt or covert) only for formal morphological reasons of checking features. Although it is possible, of course, to introduce some arbitrary feature that justifies QR, this goes against the spirit of the program, since there is no morphological evidence for such features. In the case of quantifier scope, this movement is motivated only by interpretation needs, and it is only witnessed at the conceptual interface. Even if QR could be somehow motivated morphologically, there is another issue of economy here: Raising, say, an object QNP, to obtain scope over the subject violates superiority. Since the other option of raising the subject exists, this, more economical (shorter), option should block the other.

Let me, therefore, pursue further the alternative view of QR proposed in Reinhart (1983, chapter 9). It rests on the well motivated assumption, in the framework of generalized quantifiers, that to interpret quantified NPs, there is no need to ever raise them. The only motivation for movement is to obtain scope wider than their c-command domain at the overt structure. But we noted, anyway, that this wider scope is the marked case, and it is harder to obtain than the c-command scope. (In the seventies, this was felt to be the case with universal quantifiers, but the theoretical decision was to ignore this difference between the availability of overt and covert scope. Later, it was found out, as noted in section 5.3., that the genuine scope interpretation of existential-cardinal NPs is also most readily available in their overt c-command domain.) It is far from obvious, therefore, that the computational system should be dramatically modified just to capture the marked cases. I proposed, instead, that the standard interpretation of quantified NPs is in-situ, namely their scope is their overt c-command domain. But QR may apply to create alternative scope construals. Scope outside the c-command domain, then, requires a special operation, which does not apply in the case of interpretation in situ. Interpretations derived by this operation then are more costly. This may explain why they are marked and harder to obtain[2].

---

[2]This statement of the underlying idea is from Reinhart (1983), p. 197-198. The markedness approach, stated in semantic terms, was proposed also by Keenan and Faltz (1978), who argue that lambda abstraction applies only to capture marked scope.

There are two problems that this line faces, one empirical, and one conceptual. The empirical problem is with quantified NPs which are complements of N, as in (1a). It was noted in Reinhart (1976) that this is the only structure which systematically goes against the generalization of overt c-command scope. The most available (perhaps even the only possible) scope construal is with the lowest QNP (inside the NP) taking widest scope. This is seen more clearly when there is further embedding, as in (1b).

1 a) Some gift to every girl arrived on Xmas eve.
b) Some gift to every girl in two countries arrived on Xmas eve.

I could only face these cases with an ad-hoc rule, and, indeed, May (1977) pointed out that these structures, which he labelled 'inverse linking', are the strongest argument for a QR view of scope. These still remain a mystery for the view of QR as a marked operation. (The problem is how these cases can be interpreted in situ to yield this result or, if QR is at work here, why does the result lack any air of markedness.[3]) So, I still have to leave this question open here.

The conceptual problem is the concept of markedness. Recall that what is at stake here is the question whether the interpretation of quantification requires obligatorily an operation like QR. On the standard QR view, QR is an obligatory operation which is assumed to be necessary (e.g. in order to create the variable bound by the Q operator), regardless of whether the final scope is isomorphic to the overt c-command domain, or not. On the alternative view, QR is not required for the interpretation of quantification, but it is only an optional operation for obtaining non compositional scope. This is a substantial theoretical decision, and the question is what evidence could be used to decide. The idea that QR is a marked operation rests on the intuition that it is harder to obtain scope outside the c-command domain. Though this intuition has found an empirical support (E.g. in Gil (1982)), it is still not fully clear what this means. In principle, there could be all kinds of performance factors that determine why one interpretation is preferred over the other, and the decisions regarding the structure of the computational system should not, normally, be based on statistical frequency, or other performance considerations. Furthermore, it has been noted over the years that, in the appropriate context, it may be very easy to get scope wider the c-command domain. A famous example is that of Hirschbuller (1982), which I mentioned in (71) - (72), repeated in (2).

2 a) An American flag was hanging in front of every building.
b) An American flag was hanging in front of two buildings.

If it is just as easy to get wide scope, as the c-command scope, and it only depends on context,

[3]But we should note that the later is also a problem to the QR view: If QR is the default options, i.e. (87) is the core instance of QR, why isn't it just as easy to get wide scope with V-complements.

it is not obvious what content could be given to the concept of markedness. Hence, there seemed to be no independent evidence that QR applies only when needed to obtain scope wider than overt c-command, and the debate concerning the status of QR seemed for years to be purely theory internal.

The first evidence that this may be an empirical, rather than a conceptual question, is provided by Fox' study of ellipsis (1994a, 1994b). Let us first look at the ellipsis problem he discusses.

Sag (1976) and Williams (1977) pointed out contrasts like (3).

3 a) A doctor will examine every patient. (Ambiguous)
b) A doctor will examine every patient, and Lucie will [ ] too. (Only narrow scope for <u>every</u>)

(3a), in isolation, is ambiguous, in the standard way between the construals with wide and narrow scope for <u>every patient</u>. The puzzle is that the ambiguity disappears in the ellipsis context of (3b). The same (3a), when it is the first conjunct of ellipsis, allows only the narrow (overt c-command) scope for <u>every patient</u>. (I.e. (3b) is true only if there is one doctor that will examine all patients.)

The account Sag and Williams offered for this fact is based on their assumption that VP ellipsis is an LF operation: An LF predicate is copied into the empty VP (at least in Williams' analysis). The predicate should be well-formed, and, specifically, it cannot contain a variable bound outside the copied VP. It would be easier to see how this works, if we use the version of QR proposed in May (1985)[4]. (4b), then, is the LF representing the wide scope of <u>every patient</u>, where it is extracted out of the VP and adjoined to the top IP. (4a) is the narrow scope construal. <u>Every patient</u> still undergoes QR, as is the standard assumption within the QR framework, but to capture its narrow scope, it is sufficient to attach it to the VP.

4 a) A doctor$_2$ [e$_2$ will [$_{VP}$ every patient$_1$ [$_{VP}$ examine e$_1$]]]
b) Every patient$_1$ [a doctor$_2$ [e$_2$ will [$_{VP}$ examine e$_1$]]
c) And Lucie will [ ] too.

The second ellipsis conjunct is generated, as in (4c), with an empty VP, into which an LF-VP should be copied from the first conjunct. If we copy the full (top) VP of (4a), the result is well formed. But the VP of (4b) contains the trace of <u>every patient</u>, which is bound outside the VP. Hence this is not an independent, well formed, predicate, so it cannot be copied. It follows, then, that only the LF (4a) allows interpretation of the ellipsis, hence in (3b) there is no ambiguity.

---

[4]I do this only in order to avoid introducing the definition of alphabetic variants that Sag and Williams assume. Other than that, the analysis is precisely the one they proposed.

Sag and Williams viewed this as a strong evidence for their LF analysis of ellipsis. However, Hirschbuller (1982) pointed out that this could not be the correct generalization, based on examples like (5). The wide scope construal of every building (different flags for different building) is clearly possible here, though it involves copying of a VP with a variable bound outside it, just as before.

5) An American flag was hanging in front of every building and a Canadian flag was too.

6) A doctor will examine every patient, and a nurse will too.

Fox points out that the same is true for sentences like (6), which differs only minimally from (3) (a nurse, instead of Lucie). So the question is what is the difference between (3) and these cases. Though there were many attempts at an answer, since Hirschbuller pointed the problem out, it remained, essentially, a mystery.

Fox' solution rests on the alternative view of ellipsis as a PF deletion developed in the minimalist program (see Chomsky and Lasnik (1993) and Tancredi (1992) for some of the details). The input of VP ellipsis, then, are two full derivations (clauses), and one of the VPs gets 'deleted', i.e. it is not spelled out phonetically. This is subject to parallelism considerations, which also may effect other PF phenomena, like deaccenting. The least we know about what counts as parallel derivations is that all LF operations (like QR) that apply to one of the conjuncts should apply also to the other. (Though many additional considerations may play a role.) Let us see, for example, how (6) is derived, under the construal of every patient with wide scope.

7 a) Every patient$_1$ [a doctor will [$_{VP}$ examine e$_1$]]
b) and Every patient$_1$ [a nurse will [$_{VP}$ examine e$_1$]] too.

Both conjuncts are derived in full, as in (7). QR has applied, independently to both. The result is, then, that the two VPs are precisely identical, and the second one need not be realized phonetically, so the PF is the string in (6). If QR does not apply in precisely the same way to both conjuncts, no ellipsis is possible, as witnessed by the fact that (6) cannot have different scope construals in the first and the second conjunct.

The question, now, is why the same is not true also for (3). For ellipsis to be possible under the wide scope construal, QR should apply in both conjuncts, as in (8). If QR applies freely, as in the standard view, this should be possible, and there is, again, no account for why this reading is impossible for the ellipsis in (3).

8 a) Every patient$_1$ [a doctor will [$_{VP}$ examine e$_1$]]
b) and Every patient$_1$ [Lucie will [$_{VP}$ examine e$_1$]]

9 a) A doctor will [$_{VP}$ examine every patient]
b) and Lucie will [$_{VP}$ examine every patient] too

But Fox argues that QR outside of the VP applies only if this is required to obtain an interpretation not available otherwise. In (7b) QR results in a different interpretation than that obtained if every patient is assigned scope inside the VP. But in the case of (8b), the reading obtained by QR is equivalent to the reading obtained in (9b) without QR, so there is no interpretative need that could motivate such a movement. Correspondingly, the only construal permitted in the second conjunct is (9b). Parallelism determines, then, that for ellipsis to be allowed, the first conjunct should have the same LF structure, hence, only (9) is the source of ellipsis in (3). Fox shows the same pattern in several other cases, where QR could not change the interpretation, (with two universal quantifiers in the second conjunct, and with negation). Ellipsis clearly shows, then, that QR does not apply arbitrarily, but only when movement is needed to derive a semantically distinct scope construal[5].

Fox argues that this result can follow from a certain view of interface economy, which was proposed by Golan (1993) and Reinhart (1993). The intuitive idea is that economy of this global type is sensitive to interface needs, which in this case are obtaining a distinct interpretation. Out of a set of possible derivations (from the same numeration), economy forces choosing the most economical way to derive a given interpretation. Let us examine this intuition, and its possible implementations.

2. Interface economy and markedness.

Golan (1993) noted that the type of economy which is at work in the case of superiority is not, in fact, the same as with other instances of shortest link. A striking difference is that (as widely observed) superiority is easily violated, depending on what appears to be contextual and semantic considerations. One such context, observed by Lasnik and Saito (1992), is given in (10). (10a) is a standard superiority violation. (the lower, rather than the higher wh-phrase has moved.) But (10b), where precisely the same happens in the embedded clause, is much better.

10 a) */? I know what who bought e?
b) Who e knows what who bought e?

[5]We should note that this is not sufficient to decide the question whether QR is obligatory, since it is still possible to assume that it applies in (89) inside the VP, as in May's analysis of (90). What will be ruled out by economy, in this case, is the longer movement of the QNP to an IP position. Since the longer movement yields an interpretation equivalent to that obtained by shorter movement, it is blocked by the latter. Indeed, Fox assumes, technically, this second option. What got established, in any case, is the concept of markedness that I turn to below: a less economical derivation is permitted only if it is required for a interface needs.

11 a) For which <x,y>, x knows what y bought
b) */?For which x, x knows for which <z,y>, y bought z.

This is so, however, only if who has matrix scope. The question is, in principle, ambiguous: who could be interpreted in the lower clause, as in (11b) (yielding an answer like Lucie knows what who bought). But on this reading, this remains a superiority violation, just as bad as (10a). It is only on reading (11a) (answered, e.g. by Max knows what Lucie bought) that the superiority effect disappears. (For convenience, I will ignore here the precise details of the interpretation of wh-in-situ that I assumed in part I.)

Golan argues that the economy strategy involved in superiority is interpretation-dependent, i.e. it determines the most economical way relative to interpretative goals. In the standard bad cases of superiority violations, the derivations with the long and the short movement yield precisely the same question. E.g. The derivation (12a), which violates superiority, results in the question (12b), which is precisely the same as (13b), obtained by the shorter derivation (13a). In this case the more economical derivation blocks the other.

12 a) *what did who buy e?
b) for which <x,y> x bought y?

13 a) Who e bought what.
b) for which <x,y> x bought y

In the problem case (10b), repeated in (14a), the derivation appears to violate economy as well, since a shorter derivation exists, as in (15a), where the c-commanding who is moved.

14 a) Who e knows what who bought e?
b) For which <x,y>, x knows what y bought

15 a) Who e knows who e bought what?
b) For which <x,z> x knows who bought z.

But in this case the questions denoted by these two derivations are not identical. With matrix scope of the wh-in-situ, (14b) asks for a value for who, while (15b) for a value for what[6]. So, if we try to ask the question (14b), there is no other, more economical derivation, that could derive this question. Hence, this is the most economical way to reach an interface goal.

[6]There are independent reasons for why this is so, say, for why (15a) cannot be interpreted as (14b): A wh-phrase in SPEC CP cannot take scope higher than this CP. Epstein (1992) argues that this too follows from economy, but in any case, this is a fact about wh-scope.

The crucial point in Golan (1993) and Reinhart (1993) is that a distinction is needed between the standard derivational economy, and interface economy. Derivational economy is absolute, and is not violable, regardless of interface needs. Relativized minimality is a clear instance. Wh-islands may be weaker than the cases of relativized minimality with A-movement, and they may vary in badness. Thus, (16) is worse than (17) even in English. (In Hebrew, (16) is fine, and (17) is out, for reasons discussed in Reinhart (1981).) But whatever status they have, it is not effected by context or interpretative needs.

16) *I wonder from whom you forgot what you got.
17) ?I wonder what you forgot from whom you got.

It is not too difficult to imagine which questions would be denoted by each of these derivations, had they been allowed. It is also clear that in each case, the given derivation is the only way to express the relevant question, based on the given numeration. Still, this does not improve the derivation. This means, then, that the status of (16) and (17), including the question why the first is worse, is determined by the computational system with no access to any interface considerations.

But there are other (fewer) instances of economy, which appear to be effected by interface considerations, of which we have examined so far only questions of the semantic interface. The specific example of superiority may turn out, eventually, not the best instance of this phenomenon, since there are too many superiority problems left open by this account[7]. But in

---

[7]The problem is that there does not seem to be one systematic account that captures all the contexts which were observed over the years to allow superiority violations. There are two famous contexts, which do not follow from the account above: First, the so called D-linking cases of Pesetsky (1987), illustrated in (ib), (which is better than (ia)). Next, as observed by Kayne (1984), adding another wh-focus to the sentence usually improves superiority violations. E.g. (ii) is better than (ia).

i a) */?What did who buy?
b) ?Which book did which man buy?

ii) ?What did who buy where?

The judgments on these two types of apparent violations are less clear than in the previous case. Possibly, they are effected also by focus considerations at the PF interface, along the lines I discuss in part III. (Though how precisely is not obvious yet.) But there are other problems. First, it is not easy to find other examples for interpretation-driven violations of superiority. Or, even if examples can be found, it is not obvious how they follow. E.g. Though I think (iiia) is substantially better than (iiib), I don't see how to explain this.

the case of QR, Fox' findings are a strong evidence that scope shifting of this type applies only when needed for interface needs, namely, obtaining an interpretation that is otherwise unobtainable.

The intuition behind interface economy is that, in actual human practice, deriving sentences is not an activity motivated just by a compulsion to check features, but speakers use their innate tools to express ideas, or reach other interface goals. When the computational system leaves room for optionality, therefore, the choice of operations may be directed at such kind of goals. A serious question, however, is how this intuition can be captured in the theory, without falling into vacuity.

The line taken in Reinhart (1993) is that considerations of this type apply at the stage of translating syntactic forms into semantic representations (It is not necessarily full semantic representations which need to be checked but some representation in which variables are introduced and bound.) The way it was stated there, if at the stage of translating a given convergent derivation D into some semantic representation, we discover that an equivalent semantic representation could be obtained by a more economical derivation D' (from the same numeration), D' blocks D. (I.e. D' blocks D, unless their translations are not equivalent.) This is a somewhat vague formulation. Fox (1994) proposes a precise formal statement of this intuition. He builds it into the definition of the reference set: The set out of which economy

---

iii a) ?Tell me in which subject who got the best grade?
b) *Tell me in which subject who got his best grade?

iv a) Tell me who got the best grade in which subject
b) Tell me who got his best grade in which subject.

((iva) is ambiguous, asking about the best grade relative to all grades of a given student, or relative to all students in a given subject. The pronoun in (ivb) disambiguates the sentence to mean only the first. (iiia) has only the second reading, and (iiib) is, accordingly, out.)

Furthermore, even the generalization illustrated in (14) has been challenged. Chomsky (1995, note 67) points out that (v) is not so good, and suggests that perhaps (14) only reflects "preference for association of likes".

v) *What determines to whom who will speak?

So far, then, we seem to have, more or less, a different story for each different example. The facts remain consistent with the central claim of Golan that superiority, unlike relativized minimality, is not an absolute constraint and it is sensitive to interface considerations (which can, in principle, be of different sorts). Still, at least for the time being, it would not be wise to base any theory on a phenomenon so poorly understood.

selects the most economical derivation includes only derivations which end up with the same interpretation, i.e. the reference set consists of pairs of a numeration and an interpretation. In case two derivations with an identical such pair are found, the more economical of these blocks the less economical. E.g. Using a scope shifting operation, like QR, is a less economical derivation than the same derivation without it. If the same interpretation is obtained also without it (as in the case of (8)) the reference set ends up including both the derivations with and without the application of QR, and the one with QR is blocked. But if the interpretations with and without QR are distinct, then the reference set for each contains only one member. Specifically, since, in this case, the reference set of the derivation with QR does not include also the derivation without it, the derivation meets economy.

A problem that this line faces, particularly in the framework of Chomksy (1995), is how precisely the set of operations subject to interface considerations is defined. Reference-set economy, in this framework, is a broad family, including all instances of the Minimal Link Condition (MLC). As we saw, apart from superiority, no other instance of this condition is sensitive to interpretation. If the definition of the reference set is changed, as above, why should the same not apply also to the other cases of 'absolute' economy?[8] This is a serious problem, since if we cannot define precisely the set of operations subject to interface economy, we face the danger of a vacuous theory, where all movement depends on our undefined feelings about meaning.

---

[8]In the first formulation of economy in the minimalist program in Chomsky (1993), it appeared that relativized minimality could not, in fact, be viewed as an instance of 'shortest link' (-MLC). Minimal (shortest) link was defined so as to select between two possible convergent derivations. In (16), e.g. assuming that the both the top and the embedded C are wh-marked with a strong feature, a wh-element has to move into both, for convergence. Hence, the only ways the derivations could converge are (16) and (17), which both violate relativized minimality. Under this formulation, therefore, some version of relativized minimality would still be needed, independently of minimal link. So, at that stage, the only real instance of minimal link was superiority. If this is so, it is possible to assume, as I did in Reinhart (1993), that minimal link is, in fact, a peripheral economy requirement, which unlike the other economy conditions, is sensitive to interface goals. But in Chomsky (1995), the wh feature is assumed to be interpretable, hence when a wh-element moves into the lower C, it is not erased, and it can move up to check the top C. This, then, is the shortest move, and neither (16) nor (17) can be derived (both violating the MLC). MLC, then, is a central economy principle.

If superiority and relativized minimality are, indeed, instances of the same economy restriction, it is not obvious why only the first is so easily violable. So this remains to be studied. Minimal link is also the only economy strategy involving choosing out of a reference set, which is a costly operation, so, in the long run, it may be an advantage if it turns out a peripheral strategy, while relativized minimality is governed by a different strategy. In any case, if we abstract away from superiority, all other instances of interface economy are independent of the MLC - a point I return to directly.

A way to avoid this problem, is building interface economy into the numeration, as outlined in Chomsky (1995). Chomsky assumes (not just for this problem) that any item "enters the numeration only if it has an effect on output". (Economy principle (76)). Suppose, now, that QR is a movement of some feature like 'QUANT'. Some functional feature must, then, be included in the numeration, to host this feature. This functional feature will be allowed into the numeration only if it has an effect on the output, namely, if the interpretation obtained is not identical to what will be obtained without this movement. Interface economy, then, determines the shape of the numeration: The underlying intuition may be that it is at the stage of choosing the 'stone blocks' that speakers pay attention to what it is they want to say. (Theoretically, this line resembles earlier views, that all aspects of meaning are determined in Deep Structure.) Other questions aside, if there are, as I believe, other instances of interface economy, they will all have to be encoded in the same way. This, in fact, is what guarantees that this solution is fully explicit and restrictive.

However, under both these views of interface economy (restricting the reference set, or restricting the numeration), it is assumed to belong directly to the computational system. If this is so, there is, in fact, no concept of markedness. When an operation like QR is needed for interpretation, it ends up indistinguishable in status from any other economical operation. Thus, we have no obvious explanation for the fact that quantifier scope outside the c-command domain was found, empirically, to be harder to obtain, and less common. The concept of markedness played an important role in traditional linguistics, and at the earlier stages of theoretical linguistics, but it was abandoned, largely because it did not seem feasible to give it any precise definition. Still, as we will see, there are other instances of interface economy, where markedness seems to play even a clearer role (particularly in the analysis of focus, in part III). Furthermore, there may be cases where syntactic encoding is more problematic than with QR (though, of course it is always possible, at a theoretical cost). While keeping the option of syntactic encoding, as a fall back position, it may be useful, therefore, to examine also other options.

An alternative line, which I would like to explore here is that this type of economy considerations apply at the interface, and not in the computational system itself. As far as the latter is concerned, a marked derivation is just a violation of economy. The interface question is when it is preferable to use, nevertheless, an uneconomical derivation. On the markedness view, QR is, strictly, an uneconomical operation, and whenever it applies, it involves a violation. E.g. (7b) and (8b), repeated, have precisely the same status of economy violations. This is what is meant by saying QR is always the marked option.

7 b) Every patient$_1$ [a nurse will [$_{VP}$ examine e$_1$]] too.
8 b) Every patient$_1$ [Lucie will [$_{VP}$ examine e$_1$]]

However, in the case of (8b), a choice of the marked derivation is the only way to satisfy an interpretative need. Using the costly derivation, then, is the optimal choice.

Under this view, just as in the alternative of encoding this calculation in the computational system, it is a mistake to search the evidence for the markedness of QR, or for the concept of markedness in general, in the realm of direct intuitions. If using the uneconomical derivation is, decisively, the only way to satisfy a certain interface need, the result may sound perfectly fine, and it is only indirectly that we can see that it is nevertheless marked, or uneconomical. I will return to the question of evidence directly.

The obvious question remains which type of economy is violable, namely, when is it permitted to select the optimal interface choice, in violation of economy. Once we decided to leave superiority out, it turns out that violable economy is not, in fact, an instance of minimal link economy (MLC), but rather, of the prohibition on operations not needed for convergence. All cases I will consider here involve some operation which is not, otherwise, required by anything in the computational system, like movement not motivated by feature checking, or unneeded stress shift, to be discussed in part III. It appears that the MLC is a real absolute constraint of the computational system, and interface economy governs only operations that were viewed before as optional.[9]

Eliminating optionality from the CS (computational system) is a major theoretical goal of current syntax, and it may appear that this view of interface economy is just allowing it back in, through the back door. However, the crucial point is that derivations obtained this way are, technically, disallowed by the CS, and if they are nevertheless found at the interface, they should have a clear mark of economy violations. The question of evidence is therefore crucial. As I mentioned, the marked nature of such derivations cannot be witnessed by introspection. If we allow this in the theory, without a clear way to identify marked derivations, we end up with an unfalsifiable theory.

In part III, I will discuss a cross-language type of evidence for markedness. But the crucial trait that all marked derivations should show is processing complexity. As we will see now, under this view of interface economy, marked derivations come at a huge computational cost. If no evidence for such a cost can be found for a given operation, then it does not fall under interface economy.

---

[9]There are crucial questions that I leave unattended here. Again, it is obvious that building optionality into the numeration by using optional features is more restrictive and faces no particular problems of defining its domain. However, it amounts to facing the problem with a list of such optional properties that can be included in the numeration for interpretative needs. If such lists are the only way to avoid the danger of vacuity, then we have no other choice, for now. But if the relevant interface generalization can be stated with some degree of precision, it may provide more insight into what, in fact, is going on.

3. Processing: the computational complexity of interface economy.

Under the markedness view of interface economy, the calculation involved in marked derivations is highly costly. It requires comparing two derivations - one violating economy, the other not - and deciding whether they end up equivalent at the interface. Only if they don't, the economy violation goes through. If this view is correct, marked derivations should be associated with a greater processing effort than economical ones, and we could also expect not to find too many instances where this is allowed. (The absolute, derivational, economy considerations of the computational system should reduce substantially the number of instances where this could happen.)

In the case of QR, computational complexity has been witnessed, so far, only indirectly, by empirical studies such as Gil's, which show that it is rare. Gil presented sentences with two quantifiers to subjects, in several unrelated languages, and checked their understanding of the sentence with content questions. The preferred reading (statistically) was, overwhelmingly, the overt scope (with no QR), though QR interpretations were also found. As I mentioned, statistical findings of this type are not, perhaps, a sufficiently decisive evidence for processing complexity, as long as other explanations for this performance fact have not been excluded[10]. It should be, in principle, possible to find a more direct processing evidence to the complexity involved in QR-based interpretations, which has not yet been attempted.

But we may note now that although we may expect not to find too many instances of interface economy, QR is, still, not the only case where an interface strategy of this type has been argued to be operative. The coreference strategy I argued for in Reinhart (1983) has precisely the same properties, although the economy requirement which is violated does not involve movement. In that case, the processing complexity involved has been directly witnessed and confirmed in studies of acquisition.

Abstracting away from the technical details (which are worked out in Grodzinsky and Reinhart (1993)), the coreference generalization (interface rule I) is that two expressions in a given LF, D, cannot corefer if, at the translation to semantic representations, we discover that an

[10]Still, the idea that QR involves a greater computational complexity than scope interpretations with no QR cannot seem very shocking. Recall that of all the options of scope interpretations discussed in detail in part I, those whose availability was most debated among linguists, were precisely the ones which necessarily require a QR operation. (In the seventies, some scholars refused to believe that in Some tourist read every guide book, every tourist can have wide scope, and in the nineties, Ruys, Beghelli, Ben Shalom and others, expressed the same feelings regarding, say, Two arrows hit three targets.) Given that in linguistics, the scholar and the object are one and the same, this could be some indirect reflection of the difficulties that such structures pose to speakers. Needless to say, though, we do not want to base our theory on such findings alone.

alternative LF, D', exists where one of these is a variable bound by the other, and the two LFs have equivalent interpretations. I.e. D' blocks coreference in D, unless they are semantically distinct.

18 a) The bear near Max touched him.
b) He touched Max.
c) He is Max Smith.

Let us assume that we want to let the pronouns in (18) denote Max, obtaining, thus, coreference. In (18a), there is no way to replace either him or Max with a variable bound by the other, assuming any of the standard conditions on variable binding at LF. So, no alternative LF, D', exists and, hence, the coreference strategy above says nothing on the matter (which means that it permits this interpretation). But in (18b) an LF, D', exists in which Max is replaced with a variable bound by he. The interpretations of the two LFs end up equivalent. (He touched Max, with he = Max is equivalent to He (\x(x touched x)), with he=Max.) So this D' blocks coreference in (18b). In (18c), it is also the case that an alternative LF, D', exists, that yields He (\x(x is x)), with he = Max. However, the two are not equivalent: D' is a tautology, while (18c) is not. For this reason, the coreference strategy still allows coreference in (18c).

The economy rationale behind this strategy, is that variable binding is a more economical means to identify referential identity of two expressions. Actual assignment of reference, at the interface, requires relating an expression to the set of entities in the discourse (model, domain or whatever). This is a rather complex procedure, although it falls outside of the computational system. In the case of variable binding, as with He (\x(x touched x)), this procedure has to apply once, identifying the value of one of the arguments. (he, in this example, or any other expression in the same position, in other examples.) In all other cases, it has to apply to each argument. E.g. in He touched Max, it has to apply both to he and to Max. If what is intended is the referential identity of these arguments, applying the same procedure twice, when we could have done it only once, is uneconomical. (18b), with he =Max, then, is an economy violation. But if this violation enables an interpretation not available with variable binding, it is allowed, as a marked derivation.

This coreference strategy got, unexpectedly, confirmed by studies of the acquisition of anaphora. Virtually all these studies found that children have problems with coreference. Specifically, that they do poorly on non-coreference, as in (18b). ((18c) was not studied). The actual statistics is particularly interesting here: Children get the right responses with the permitted coreference case of (18a) at about 100%. But on non-coreference, like (18b), we find about 50% of correct responses. This means that they guess on the second. If children did not know the coreference rules at all, we should not expect to find this specific guessing pattern. Not knowing a rule of non-coreference exists, children should always allow it, i.e. we should get about 100% of allowing coreference for both (18a) and (18b) (or, at most, a guessing pattern on both).

We should note that, although, this is one of the most studied areas of acquisition, no account has been actually offered for these performance facts. The most elaborate attempt at an account is Chien and Wexler's (1990, 1993). They assume that to capture these facts, an extension of the standard binding theory is needed. Conditions A, B, and C, govern coindexation, or contraindexation, but in addition, we need another condition P, which states that contraindexing means non coreference, unless the contexts permits coreference. Next, they assume a theory of maturation. So, it so happens that conditions A, B and C mature before condition P. Hence, children know how to put indices correctly on their expressions, but what they don't know is what these indices mean, since they can't yet make use of condition P. So, this is the source of the impression that they don't know conditions B and C. Although this may appear, at first sight, a solidly unfalsifiable theory, it is interesting to observe that, nevertheless, it cannot explain the facts it was designed for. If children don't know condition P, namely they don't know that the contraindexing in (18b) entails non coreference, how do we get the 50% pattern? They should just allow coreference in these cases, or at least, guess also on (18a).

As explained in detail in Grodzinsky and Reinhart (1993), these results follow directly if the interface coreference strategy (rule I) is innate, like anything else, and no maturation is assumed. When processing (18b), Children know that if they allow coreference, this is an economy violation, and they know exactly what they have to do to decide: construct an alternative LF, and hold the two LFs, while checking if their interpretations are equivalent in context. But they lack the computational ability (storing and processing) to go through with this, which is known to develop with age. So having started, and gotten stuck, they give up and volunteer a guess. The processing of (18a), by contrast, is much easier. Since no alternative derivation is found to begin with, the task of approving coreference can be completed without comparing two semantic representations.

In sum, then, Violating economy to meet an interface need, does not come for free. It requires a heavy processing cost, which is what is meant by calling such derivations marked. In view of this complexity, we would not expect to find instances of such strategies all over the place. At the same time, since interface economy is a general strategy, ruling out optional operations when not needed for interface needs, the type of operation itself is not determined by this strategy, and instances can be found in unrelated modules. Indeed, (abstracting away from superiority), we noted so far only two instances, but with two independent types of operations: movement operations (QR) and anaphora resolution, namely, the procedures relating expressions to the discourse domain. (rule I). In part III, I will discuss one more instance in another independent type of operation - the stress rule of PF.

We should keep in mind that this type of interface economy, which governs superfluous operations and which entails markedness, is just one of the interface strategies. It is only if evidence for the computational complexity is found for some operation, that we can conclude that it falls under this type. Operations for which no such evidence can be shown, belong elsewhere. One option noted already in section 2, is that they are directly encoded in the CS,

say, as optional features whose selection is governed by the economy requirement on the numeration set, stated in Chomsky (1995). Or they may be governed by different strategies which apply only at the interface. One such unrelated strategy, is the assessment procedure, sensitive to topichood, which I discuss in part IV.

Appendix: A speculation

I would like to point out one potential derivative of interface-economy view of QR, though at the present, this is only a speculation. (This implication is independent of whether QR is implemented as an economy violation, or as an optional feature subject to numeration economy.)

As is well known, none of the existing accounts for weak crossover gets close enough to capture the facts. Historically, one of the reasons why this got to be a problem is, precisely, the introduction of QR. If quantifiers cannot have scope outside their c-command domain, nothing is needed to explain why (19) does not allow the pronoun to be bound. Trivially, variables can be bound only in the scope of the binding operator.

19) *His$_i$ mother praised every speaker$_i$

But since it was assumed that every speaker undergoes QR, the pronoun ends up in its scope, hence the problem in (19) was understood as an anaphora problem specific to natural language, that cannot follow from just standard logical considerations.[11] But all attempts to capture the distribution of variable binding as an independent syntactic issue have, so far, run into substantial difficulties.

Ruys (1992) argued, in view of these difficulties, that we should reconsider the assumption that, in fact, the availability of pronouns as bound variables is dependent on the scope of the operators that bind them, and nothing else. However, executing this under the assumption of obligatory QR, which he attempted to keep, required a substantial enrichment of the machinery, leading to a certain degree of arbitrariness.

Possibly, the present view of QR, subject to interface economy, may open the way to reexamine our basic assumptions. Support comes from comparing (19) to cases like (20).

[11]This was, essentially, the argument in Chomsky (1976). The same conclusion holds also for approaches assuming that QR is only optional. Thus, I argued in Reinhart (1983) that the fact that the pronoun can never be bound in that domain, regardless of whether the marked QR operation applies or not indicates that variable binding and scope are independent.

20) a) A copy of his$_i$ speech was placed in front of every speaker$_i$
b) ?A friend of his$_i$ mother praised every speaker$_i$
c) ?Someone paid by his$_i$ mother praised every speaker$_i$

21) ?/*Every copy of his$_i$ speech was placed in front of every speaker$_i$

If we assume that QR takes place only when needed to yield a distinguishable interpretation, there is no need for it to apply in (19) Hence, <u>every speaker</u> does not have scope over the pronoun, and it cannot bind it. But in (20), the derivations with and without QR yield distinguishable interpretations. Only if <u>every speaker</u> raises, it gets scope over the subject. If it does, it has the pronoun in its scope and can bind it, in apparent violation of all standard weak crossover requirements. (In the specific example (20a), the reading obtained by QR happens to be more plausible than the one without it. This may be the reason why it is easier to perceive this reading, and consequently, to get anaphora, than in the (b) and (c) cases). Another entailment of this line is that in (21) anaphora is much worse than in (20), since QR is not motivated by any added interpretation.

Needless to say, there are many questions to answer, before we can conclude that this is the right road. Not least is the question how we rule out the possibility that binding a variable is, itself, a sufficient license for QR.

III. FOCUS - THE PF INTERFACE[12].

There is a certain resemblance in the history of the view of quantifier scope, and of focus, in theoretical linguistics. At the earlier stages, e.g. Chomsky (1971), focus was viewed, essentially, as a property defined on PF structures. The basic idea was that sentence stress is assigned independently, by the phonological rules, and the interface systems make use of this available stress in relating a sentence to its context, namely, signalling the focus and presupposition structure. The focus was defined as any constituent containing the intonation center of the sentence. The problem that this view encountered was that it rests on the notion of 'normal intonation'. Specifically, a distinction was needed between this type of normal stress, and more marked stress options required by discourse needs[13]. This distinction has been often challenged since, and a common claim in focus studies is that it is impossible to distinguish between 'normal' and 'marked' intonation patterns. Hence, focus intonation cannot be assigned at PF independently of the semantics of the sentence, and it must be the other way around: sentence intonation reflects its independently determined focus structure. This seems to have been the winning hypothesis for years. In fact, in Chomsky (1976), where QR was introduced for the questions of quantifier scope (and its interaction with anaphora), the view is that focus is signalled by QR, and, consequently, any constituent that can be raised by QR can serve as focus.

However, in a pioneering study, Cinque (1993) reopened this issue, and proposed, in essence, returning to the PF view of focus. There are substantial conceptual advantages to this line, which we will see in detail. But, as before, it requires a distinction between normal ('neutral' - in his terms) and marked patterns of sentence stress. It may appear, therefore, that his analysis could face the same problems for which it was abandoned in the previous round. So, my major concern here is, first to provide further support for Cinque's approach, and next to explore further the concept of markedness. As in the case of QR, I view a marked derivation, as an economy violation. One of the reasons the notion seemed problematic is that it cannot be directly witnessed by introspection: When the marked option is the only way to reach an interface need, it sounds perfectly normal. It is only when a more economical option is available, that the marked nature of the alternative is visible. In the case of focus, evidence for markedness can come from cross-language examination. Zubizarreta (1994) shows that if a

---

[12]The analysis of the Dutch scrambling in this part is based on work together with Ad Neeleman. I would also like to thank Hubert Haider and Eric Reuland for many helpful comments.

[13]"Special.. processes of a poorly understood sort may apply in the generation of sentences, marking certain items as bearing specific expressive or contrastive features that will shift the intonation center." (Chomsky, 1971 p. 199)

language has the means to get a certain focus structure without applying the marked operation (say, by choosing an alternative permissible derivation), then its application yields visibly bad results in that language.

As a case-study of this concept of focus and markedness, I will examine object scrambling in Dutch.

1. Object scrambling in Dutch

A VP containing V, O, and ADV (or PP) can be realized in Dutch in the two ways illustrated in (1).

```
1     a)          /\              b)'scrambling':  /\ or   /\
            ADV                                  O          O
                O                                    ADV        ADV
                    V                                    V          t
                                                                      V
```

Much theoretical debate has been generated around the question how the scrambling structure (1b) is generated. The two major positions are that it is derived by A (or A') movement (as in the right hand tree), and the alternative, proposed in Neeleman (1994), that it is base generated, just like (1a), the only difference being the adjunction cite of the Adverb. Although the second seems to me empirically and conceptually advantageous, for the present discussion, it is not necessary to decide between them, and I will keep the name 'scrambling' for the structure in (1b), just for convenience.

This, apparently minute, structural difference correlates in Dutch (and German) with substantial differences in acceptability, and, consequently, has led to substantial theoretical machinery devised to explain these differences.

It has been observed that there appear to be definiteness effects associated with scrambling. Compare, e.g. (2) and (3). (Examples and their judgments are from de Hoop (1992, p. 138-139), hereafter dH).

2 a) dat de politie gisteren taalkundingen opgepakt heeft
that the police yesterday linguists arrested has

b) */?dat de politie taalkundingen gisteren opgepakt heeft

3) dat de politie de taalkundingen gisteren opgepakt heeft
that the police the linguists yesterday arrested has

The indefinite linguists seems fine in the standard order in (2a), but in the scrambled version in (2b) it is worse. On the other hand, the definite the linguists in (3) is fine in the scrambled position.

In fact, however, it is not only definite NPs that can scramble. It has also been noted (attributed to Eric Reuland) that indefinites can scramble under their 'specific' use, as in (4b), or under their generic use. In (4a), e.g. the change of the adverb to always allows the same linguists of (2a) to scramble, since a generic interpretation is available for them.

4 a) dat de politie taalkundingen altijd opgepakt heeft
that the police linguists always arrested has (dH p. 138-139)

b) dat de politie een kraker gisteren opgepakt heeft
that the police a squatter yesterday arrested (dH p 50)

The popular approaches to this problem attempt to derive such facts from the properties of indefinite NPs. I.e. there should be something in the indefinite NP itself, or its relation to the context that determines whether it can 'scramble'. They follow the assumption that familiarity, d-linking, specificity, etc, are somehow encoded syntactically. Diesing (1992) and de Hoop (1992) argue that all indefinites are always ambiguous between what they call the 'weak' and the 'strong' reading of indefinites. For Diesing, 'strong' means 'presuppositional'. For de Hoop, an indefinite is 'strong' iff it is either a. specific- referential, b. generic, or c. partitive. (She assumes that cardinals like two cats are always ambiguous between the existential reading, and a reading like the partitive two of the cats.) Both introduce syntactic machinery that distinguishes the two cases. I will discuss Diesing's machinery in part iv. de Hoop assumes that the two types of indefinites get different types of case. Her major evidence for this distinction are the scrambling structures we are considering, so it may be useful to view her analysis in some detail.

The first generalization which de Hoop proposes is given in (5). This captures both the cases of definite NPs, and her 'strong' indefinites.

5) de Hoop's generalization: Only strong NPs can 'scramble'.

To derive this generalization, de Hoop postulates that weak case can only be licensed in the original Deep Structure position of the object, while strong case is freer. (A generalization similar to (5) under different stipulations is assumed also by Diesing.)

Several problems with this generalization can be mentioned, but the real serious one, which de Hoop notices herself, is that it does not seem a sufficient condition[14]. In (6b) ((based on

[14] In fact, it is not clear that 'strength' is even a necessary condition. Eddy Ruys noted the following:

observations by Kratzer), the indefinite is generic, still it does not scramble and only the unscrambled version (6a) is allowed. This can be compared to the minimally different (7), which allows scrambling. The same point is illustrated in (8), from Ad Neeleman. ((8b) shows that this is not due to the fact that the phrase take a bus is idiomatic, since wh-extraction is possible.)

6 a) omdat ik altijd een kat heb
because I a cat always have
b) *omdat ik een kat altijd heb (dH 72, p. 163)

7) omdat ik een kat altijd liefheb (dH 71, p. 163)
because I a cat always love

8 a) *Dat ik een bus altijd neem
that I a bus always take
(b Welke bus neem jij altijd t
which bus take you always)

There is also an important difference between these cases and the standard definiteness effects illustrated in (2) to (4). The judgments on the latter are known to be subtle. There is lots of disagreement, and usually it is easy to find contexts which make them perfect. But the judgments in (6)-(8) are solid and clear. So, so far, the 'strength' generalization (5) explains the less clear cases, while leaving out the clear ones.

Attempting an account for (6)- (8), de Hoop discovered the important descriptive generalization (9), which has not been noted before.

9 A descriptive generalization:
"In Dutch, scrambling of the object yields the same semantic effect" as the contrastive predicates with stressed verbs in English. (dH, p. 165)

---

i a) dat elke arts wel een of andere ziekte (meestal) met plezier behandede.
That every doctor some disease or other (usually) happily treated.

b) dat elke arts een ander ziekte (altijd) met penicilline behandelde.
that every doctor another disease (always) with penicillin treated.

Under any definition, it is extremely difficult to view the indefinites some disease or other in (iia) as 'specific'. In (ib), it is clear that the indefinite another disease does not even have wide scope: It is dependent on the universal, which is, typically, a 'weak' reading.

For illustration, given a sentence like (4b), repeated in (10). Its unscrambled version corresponds to (11a), with normal sentence intonation, but the scrambled version in (10), is best translated into English as (11b), with bold letters standing for heavy stress.

10) dat de politie een kraker gisteren opgepakt heeft
that the police a squatter yesterday arrested.

11 a) The police arrested a squatter yesterday. (unscrambled)
b) The police **arrested** a squatter yesterday. (scrambled)

This means that the appropriate English translation of (6b) and (7) are as given in (12). Indeed, (12a) is weird in English, as its Dutch counter part with scrambling.

12 a) #because I always **have** a cat
b) because I always **love** a cat.

So, the basic idea is that when contrastiveness is impossible, we should also not be able to get scrambling. However, de Hoop now attempts to derive the descriptive generalization (9), while still assuming the condition (5) that only strong NPs can scramble. For this she introduces the following theoretical account, and a new principle, the POC.

13 The theoretical account:
"If an object receives a strong reading, predication needs to be contrastive." "This principle holds more generally for all NPs of type <<e,t>,t>" (i.e. generalized quantifiers) (dH p. 165)

<u>Principle of contrastiveness (POC):</u>
A strong NP needs a contrastive predicate. (dH p. 168, stated formally in (80), p. 166).

While the descriptive generalization in (9) is important and correct, it is hard to see what exactly could be meant by the generalization of the POC, from which it is supposed to be derived. As stated, it wrongly entails that, universally, we cannot find a strong NP with a non contrastive verb or predicate. E.g. that (14a,c) are ill formed, and only (14b) is allowed in English.

14 a) Max read every book. (normal stress on <u>book</u>)
b) Max READ every book.
c) Max has already read every book.

Let us, therefore, leave the POC aside for the time being, and adopt the descriptive generalization (9). Now, let us review the new theoretical machinery which we accumulated, apart from (9), to handle this apparently minute structural variation in Dutch. The assumptions

are: a. Indefinites are always syntactically and semantically ambiguous. b. There are two kinds of case an indefinite object can get: weak and strong (partitive). c. Weak case can be assigned only in DS position, and does not allow A movement. These three assumptions combined, still only capture the very weak indefiniteness judgments of (2)-(4). The stronger judgments are captured only by the contrastivness generalization (9). An obvious hope we could entertain now is that, perhaps, something like (9) should be sufficient, alone, to explain all cases, and next that (9) could, indeed, be derived from independently established principles.

The point of departure which the approaches discussed here share, is that whatever it is that determines scrambling options of NPs should be searched inside these NPs. Namely, that if scrambling is movement, this operation is motivated by properties of the moved NP. If this is so, it is surprising to discover that the verb must also have some special properties, like contrastivness. The next move is to check, instead, the effects of scrambling on the downstairs, e.g. on the verb.

A fact that we will soon see in more detail is that in the unscrambled version, main sentence stress falls on the object, while in the scrambled version it falls on the verb. A growing attention is paid recently to the option that certain types of movement (or other structural choices, like adjunction) are motivated by PF (phonological or prosodic) reasons. These PF considerations, on their part, may interact with the focus structure of the sentence. This was proposed in depth by Zubizarreta (1994), who shows this for a variety of structures in Romance[15]. Zubizarreta's line is based on and develops the pioneering research of Cinque on sentence stress, to which we turn now.

2. Cinque's theory of stress and focus

2.1 Sentence stress.

The broader issue Cinque (1993) is concerned with is phrase and compound stress, but the instance of this problem which is relevant for us here is sentence-stress. Previous analyses, which followed, in various ways, the nuclear stress rule of Chomsky and Halle (1968) assumed that this rule is parameterized, to capture the stress patterns across languages. Halle and Vergnaud (1987) developed a metrical approach to this rule (following the metrical-lines analysis of word stress, as first proposed by Liberman): The basic idea is that the nuclear stress rule (NSR) applies cyclically, where the cycles are determined by syntactic constituency. The input of the procedure is the sequence of (non compound) word's stress, marked by asterisks, and represented as a line. A new line is introduced for each new cycle. The NSR, then, locates the prominent stress of this line. My summary of how this works will be simplified. A more

[15]That Zubizarreta's approach may be useful for the analysis of scrambling in Dutch was proposed also, briefly, in Delfitto and D'hulst (1994).

detailed summary can be found in Cinque (1993). For illustration, let us check how we derive the fact that main stress in the simple sentence (15a), falls on book. Throughout, I will represent the word carrying the main stress of a sentence with bold-face (for reasons of typographic visibility).

15 a) I read the **book**.
b) (Dat) I het **boek** las (Dutch)

16) [I [read [the book]]]
a) Line 1 (=word line 3): [* [* [*]]]
b) line 2 (VP cycle): [[*]]
b) line 3 (IP cycle): [*]

The output of word-stress for (15a) is (16a) (which is assumed to be metrical line 3). NSR then selects one of the word-stresses of line 1, and places it in line 2. The same holds for line 3.. Of course, the question is how the rule knows which asterisk to place on the next line. (A simple idea such as 'take the rightmost asterisks' won't do, e.g. for the Dutch equivalent (15b).) Halle and Vergnaud (H&V) first define the cycle as a syntactic constituent containing at least two asterisks (stressed words). In this case, one is defined as the head of the constituent line. Once the head is identified, the NSR proceeds to project the head of each line into the next line. The gist of this procedure is, then, stated in (17).

17) Nuclear stress rule
Locate the heads of line N constituents on line N+1.
18) Parameter setting for English (on line N(N>/3): [..+HT, right]

But the crucial question, now, is how we identify the relevant head for (17). This is why (17) has to be parametrized. For English, as the parameter is set in (18), the head must be in terminal position of its constituent, and this position is to the right, thus book is selected as the relevant VP asterisk. Given all these assumptions, the derivation in (16) goes through, giving the right result for English. It is not a trivial matter to define the parameters for a "mixed" language like Dutch. If we define it as left-headed, we will get correctly the stress of (15b), since the VP leftmost stress (boek) will be projected. But with an intransitive sentence, the leftmost stress will be the subject, which may then get the main stress, incorrectly.

Cinque's insight is that, in fact, no parametrization of the stress rule is needed. Apart from the empirical problems of such parametrization, it is doing nothing more than an unneeded duplication of the mechanism which governs, independently, word order variations in syntax. Assuming that we need independently to know what is the direction of recursion in a language, the same (and better) results will be obtained with applying the one universal stress rule, starting with the most embedded constituent of the sentence.

The basic idea is as follows: Let us assume that the first cycle of the stress rule is the most

deeply embedded stress, i.e. a category containing only one (word-level) stress. The stress rule now needs no mention to heads or their order, and it can be stated with a slight simplification in (19). As far as I can see, the rest follows with no further assumptions. I should mention that I am not fully loyal to Cinque's actual execution. - He assumes a greater machinery than I do here, though I think I capture correctly his intuition. Nothing here hinges on this being the case, and if my presentation is mistaken, one can go back to Cinque's precise formulation.[16] Let us see how the derivation of the stress of (15a) follows.

19) <u>Generalized stress rule</u>
Locate the stress (asterisk) of line N on line N+1.

20) [Max [read [the book]]]

a)	Line 1 (=word line 3):	[* [* [*]]]
b)	Line 2 (NP cycle):	[[[*]]]
d)	line 3 (VP cycle):	[[*]]
c)	line 4 (IP cycle):	[*]

Let us assume that the most deeply embedded constituent is the object (a point I return to in a second). The first cycle-line, (20b), is then the NP (or N). Since there is only one stress for this cycle in the previous line, it is this stress which projects to the present line. From then on, there are no more options, and each cycle projects this same stress.

Thus, the gist of the analysis is that the main stress of the sentence will always be on its most embedded constituent, namely, on the node we started stress-processing with. Of course, Everything depends now on the correct identification of the most embedded node. Specifically, the problem arises in the case of sisters (both carrying stress). Cinque argues that the answer lies in the order of recursion. Given two sisters, the most embedded one is that occurring on the recursive side of the tree. At first sight, this may seem like begging the question, but Cinque's point is that the order of recursion, or whatever determines word order is a problem independent of stress, the answer to which is the goal of current syntax. Once the answer is found, the stress pattern should follow. Thus, in a right branching language like English, in the VO structure, in (21), the most embedded node is the object. In a left-

[16]Cinque's stress rule (10), P. 244) still includes the formulation in (3), which assumes heads. It includes also an additional requirement that an asterisk on line N must correspond to an asterisk on line N-1. In his actual analysis, he starts with the next XP cycle (e.g. VP), just like H&V. But curiously, he omitted the requirement that the cycle contains at least two asterisks, and he adds that "this simplification is crucial to obtain the correct results" (footnote 7, p. 244). Indeed, this omission enables the analysis to work also without the previous assumptions, which is why I think this is what he actually intended. In any case, I do not think that there is anything at stake here apart from whether the machinery can be reduced. And I assume that the way I presented Cinque's analysis is precisely equivalent, empirically, to his.

branching language, like Dutch (in this relevant structure), it is again the object.

21) **Asymmetry of sisters.**

English		Dutch		
V'		V'		
V	O	O	V	
	*	*		

Zubizarreta (1994) argues that, in fact, it is not correct to talk about just order of recursion here, and depth of embedding is determined by head- complement relations.

With this assumed, the Dutch 15b, repeated in (22) is derived as in (23).

22) (Dat) ik het **boek** las /I the book read.

			[ik [[het boek] las]]
23	a)	Word stress:	[* [[*] *]]
	b)	NP cycle:	[[[*]]]
	c)	VP cycle:	[[*]]
	d)	IP cycle:	[*]]

The intransitive case appears non-problematic, at this stage: Given a sentence like (dat) ik **las**/ I read , the first cycle assigns stress to V (or to VP - nothing hinges on this, in this case). Since the VP and the subject are not sisters, the issue of embedding does not arise, and it is clear where the stress-processing starts. Hence, the main stress will fall on the verb.

More problematic are structures where the subject (or another adjunct or specifier) is a complex constituent, containing more embedding than the VP. The main stress still falls in this case on the deepest constituent of the VP, and the question is how this happens. Cinque assumes that the subject constitutes a cycle of its own. In this, he follows Halle and Vergnaud, who noted, independently of this problem, that the subject always gets secondary stress (higher than non-stressed nodes in the VP). The issue, then, becomes that of how to merge two cycles, each carrying its own main stress. Cinque defines, for that, the notions of major and minor paths of embedding. The main stress always falls on the major path, but when a minor path joins it, it gets a secondary stress (one asterisk). Zubizarreta (1994) offers a different formulation of this merging, sensitive to the complement/adjunct distinction, but for our purpose here these details are not crucial.

Cinque argues that his stress rule applies directly to syntactic constituents and no notions like a phonological or prosodic phrase are needed. The question what are the relevant constituents for phrasal stress has been a subject to much debate. Cinque's line contrasts with the view

developed by Selkirk (1984), where it applies to phonological phrases, related, but not isomorphic to syntactic constituents. If Cinque's analysis can be maintained, it is clearly advantageous, being the more minimal one. In any case, Zubizarreta (1994) points out that Cinque's analysis can also be stated to apply to phonological constituents.[17]

2.2 Main stress and Focus.

The analysis of sentence stress outlined so far is independent of any discourse considerations: It is impossible to utter a sentence with no prominent stress, so the PF rule we examined -(19) - determines where this stress will fall. The main stress of the sentence, which is assigned by this rule, is just a particular instance of stress assignment which is needed independently (e.g. for units smaller than a sentence). However, sentence accent interfaces with the theory of discourse, via the notion of focus. Focus, which is roughly viewed as the most informative part of an utterance, is usually identified by prominent stress. The gist of Cinque's proposal is that the set of possible (neutral) foci in a sentence is determined by its main stress, i.e. by the same rule of phrasal stress. - I return directly to how precisely this works. On this issue, of the relations between main sentence-stress and focus, there exist two conflicting positions: The one that Cinque returns to is that possible focus selections are restricted by an independent PF stress rule, and the other is that there is no such thing as a (neutral) PF stress, and the main stress of the sentence is determined solely by its relations to discourse, i.e. by focus. Cinque surveys common counter arguments to the position he defends and concludes that discourse considerations may at times interfere with the results of the phrase-stress rule, assigning a different stress-prominence. But he assumes that the two types of prominence can be distinguished. For him, the relevant distinction is that between sentence grammar and discourse grammar. The later can change the output of the computational system: If in a given context, it is appropriate to use as a focus a constituent which was not assigned the main stress by 'sentence grammar', 'discourse grammar' assigns an additional stress to this constituent, or destresses the original prominent stress.

Zubizarreta (1994) develops this line, and argues that the relevant distinction is that between a neutral focus and a marked one. Neutral focus intonation is often characterized as that intonation under which a sentence could be uttered 'out of the blue', namely, the whole sentence is asserted (as "new") and none of its constituents need to be pre-assumed in the context (no "presupposition"). Zubizarreta argues, then, that what Cinque's stress rule determines is the neutral focus intonation of a sentence. When a sentence with this intonation

[17]Zubizarreta points out (footnote 14) that the question whether the prosodic phrase is determined semantically or syntactically does not have much empirical content. Selkirk argues that the intonational phrase must form a sense unit, where two constituents constitute a sense unit if they stand in a modifier of-head or an argument of-head relation. But the notions assumed in this definition: modifier argument and head are, anyway, syntactic notions.

is uttered 'out of the blue', the full sentence can be viewed as the focus phrase.[18] But the central point of Cinque's and Zubizarreta's analysis is that, under the same neutral-focus intonation, a sentence can be used also with only one of its constituents as the focus (and the rest pre-assumed). Crucially, the full set of the possible (neutral) focus constituents of the sentence is determined by the same rule of phrasal stress. Cinque's generalization is given in (24), and we may label it the focus rule.

24) The focus rule
The focus of IP is a(ny) constituent containing the main stress of IP, as determined by the stress-rule. (This is Cinque's 'Sentence-grammar' focus, and Zubizarreta's 'neutral focus')

To see what is the set of possible foci generated by this rule, let us look at the sentence (25), whose main (neutral) stress falls, as predicted on the object, a desk. This stress is determined, as we saw, cyclically, by assigning each new cycle the main stress of the previous one. There are 3 cycles: the NP, the VP, and IP, and each of them has the same main stress. Each of them, then, can be said to carry the main stress of the sentence.

25 [My neighbor [is building [a desk]]]
* * *

a) NP cycle: [*]
b) VP cycle: [*]
c) IP cycle: [*]

The focus rule (24) now determines that each of these constituents can serve as the focus. This means that with this main stress, the sentence can be uttered in contexts in which it is appropriate for any of these three constituents to serve as focus. This is illustrated in (26). The notation I will use throughout is **bold-face** to mark the word which carries the main stress, and underlining for the constituent which is the focus selected in the given context.

26 a) -What's this noise?
-My neighbor is building a **desk**.
b) -What's your neighbor doing?
-My neighbor is building a **desk**.
c) -What's your neighbor building?
-My neighbor is building a **desk**.

[18]This is assumed under different wordings since Chomsky (1971) and Jackendoff (1972), but has recently gained more attention in work by Vallduvi, Engdahl, and Herman Hendricks.

d) -Has your neighbor bought a desk already?
#-My neighbor is building a **desk**.

e) -Who is building a desk?
#-My neighbor is building a **desk**

In (26a) we have an instance of 'out of the blue' context. Here the option (25a) is selected in the answer, with the whole IP as focus. (26b,c) illustrate contexts for the selection of (25b,c), respectively. The crucial point is that in all three contexts precisely the same main stress is used. But the same main stress cannot be used in (26d,e). In (26d), the context determines that the relevant focus should be only the verb. But the verb is not one of the constituents that (24) selects as possible foci for this structure, since it does not, itself carry the main stress. The same is true for (25e), where the context forces the selection of the subject my neighbor as the focus.

As Cinque notes, his analysis goes back, in its essence, to the view of focus in Chomsky (1971). Another way to check the prediction that any of the constituents dominating the main (neutral) stress can serve as focus, is checking the set of possible substitutions. E.g. in the context of a yes/no question in (27), modelled after Chomsky's example, the different answers correspond to different selections of focus in the question. The focus constituent in each answer, which is underlined, substitutes one of the possible foci in the question, namely one of the constituents dominating the main stress of the question.

27 Are you [looking for [a passenger with [a red [**shirt**]]]]?
a) No, I am looking for a passenger with a red **tie**
b) No, I am looking for a passenger with a **coat**
c) No, I am looking for a member of he **crew**
d) No, I am just wandering **around**

The basic idea, then, is that the main stress assigned by PF enables a sentence to be used in a variety of contexts, since it permits a large set of possible foci, from which the context can select the appropriate one. Nevertheless, there may be contexts requiring a constituent not in this set to serve as the focus. Sticking to the structure SVO, in English or SOV in Dutch or German, the focus set defined by (24) is (28b).

28 a) [ip S [vp V **O**] / [ip S [vp **O** V]
b) Focus set: {IP, VP, O}

Constituents not included in the focus set in both structures are V and S. I.e. sentences leaving PF with the standard main stress cannot be used with their subject or verb as the (only) focus. That this is indeed so, was witnessed by the inappropriateness of (26d,e), repeated in (30a) and (31a). For such contexts, Cinque's discourse-rule will have to apply. Ignoring certain details of Cinque's actual procedure, this rule may be stated as (29).

29) Marked focus rule
Relocate the main stress on a constituent you want to focus.

With this rule applied, the sentence can be used in the context of (30, 31), as illustrated in their (b) answer.

30 -Has your neighbor bought a desk already?
a) #-My neighbor is building a **desk**.
b) -My neighbor is **building** a desk.

31 -Who is building a desk?
a) #-My neighbor is building a **desk**.
b) -My **neighbor** is building a desk.

Under this analysis, then, the focus use in (30b-31b) is viewed as marked, since it is obtained by a special operation undoing the results of sentence stress. This, in a way, is the heart of the analysis, and the center of the debate concerning focus and sentence stress. The idea that a systematic distinction can be drawn between marked and neutral stress at the sentence level has been often challenged (with the alternative view being that stress at this level is determined by focus, and not conversely.) The crucial question is whether an appropriate definition, and further supporting evidence, can be found for this distinction. I will address this issue in detail in section 4, but let us assume, for now, that this distinction can be maintained, as I will argue, and look at some of its consequences.

On this view, using marked stress is costly and uneconomical, involving an additional operation. We would expect that this would be done only for a good reason, namely, when there is no other way to express the intended focus relations. English, with a rather restricted word order does not have too many choices here. But languages with more word order options, may find ways to express more focus-structures with neutral stress. Cinque compared the following sentences in English and Italian:

32 a) Johnson **died**
b) **Johnson** died

33 a) Johnson e' **morto**
b) E' morto **Johnson**
c) #**Johnson** e' morto.

In English, to create a focus structure with focus on the subject, one must use the marked stress rule, to yield (32b). In Italian, there is an option of raising the subject, as in (33a), or not, as in (33b). In the first case, neutral stress will fall on the verb, as the most embedded constituent. In the second, it will fall on the subject. Thus, Italian allows expressing both focus structures

of the English (32) with no appeal to marked stress. Next, Cinque observes that the use of marked stress on the subject, as in (33c), is inappropriate (even in the right context, which he provides, following Schmerling). This is so, since there is no reason for this option- It does not give us any option that could not be obtained with an alternative derivation with a neutral stress. Another way to describe the inappropriateness of (33c) is that the function of subject raising is precisely exempting the subject from the focus role which the main stress forces on it, in embedded position. (This is consistent with the observation, analyzed in depth in Pinto (1994), that when the subject is D-linked its movement strongly preferred. - D-linked constituents are not particularly happy foci.) Hence it appears self-defeating to then apply a special marked operation to give this stress back to the subject. I will return to the type of economy calculation which underlies this informal description in section 4.

This idea is taken much further in Zubizarreta's pioneering research on the relation of focus and movement. Her generalization, based on an extensive study of Romance, is that movement out of VP may be due to phonological reasons - namely, to change the stress pattern, hence the focus structure of a sentence.

With this assumed, let us go back to the analysis of object-scrambling in Dutch.

3. A focus account for object-scrambling

As we saw already, in the discussion of (22), when V' in dutch contains V and O (or another complement), the main sentence stress falls on O (in the standard SOV order). When it contains only V, it falls on V. We may note that the predictions of Cinque's analysis were confirmed, independently in Gussenhoven's 1984 study of Dutch stress. Let us see one further illustration, originally noted by Gussenhoven, and discussed in Neeleman (1994).

34) Dat ik [v' op een **bankje** wacht]
that I on a bench wait
that I am waiting for a bench.

35) Dat ik [v'op een bankje [v' **wacht**]]
that I on a bench wait
that I am waiting on a bench

In (34), the PP is a complement. As predicted, stress falls, then, on this PP, i.e. on its most embedded node which is the NP complement. In (35) - an apparently identical string - the PP is a locative, hence an adjunct. The most embedded node, then, turns out to be the verb, which gets the stress. Thus, these two stress patterns correlate with different meanings and different structures.

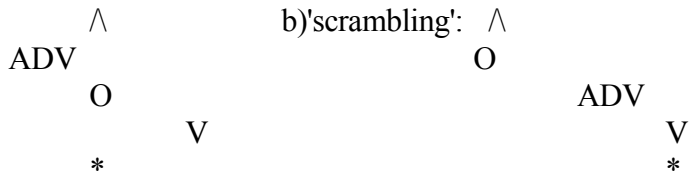
Turning now to scrambling structures, the scrambled object is not in a complement position, but it is higher than V'. Hence, the most embedded node in this structure, is the verb, just like in (35b), or in sentences with intransitive verbs. Indeed, we see below that the scrambled sentences in (b) have a different stress pattern than their non-scrambled counterparts in (a), with neutral sentence stress shifting to the verb. It does not matter, in this regard, whether the object is definite, as in (36), or indefinite. What determines the stress pattern is the fact that the object is not a sister of V, in the scrambled version[19].

36 a) Dat ik gisteren het **boek** las
b) Dat ik het boek gisteren **las**
that I read the book yesterday

37 a) Dat ik altijd een **brief** verscheur
b) Dat ik een brief altijd **verscheur**
that I always tear up a book

Let us see, now, what difference in focus options is entailed by this stress system, given the Cinque-line of focus analysis. The stress in both structures, summarized in (38), is neutral stress. Hence, it determines the focus set in the way we observed in the previous section.

38 a) b)'scrambling':



Focus set: {IP, VP, O} {IP, VP, V}

The difference is that in the 'base' structure (38a), the object is included in the focus set, but the verb alone is not. In (38b), on the other hand, the verb is in the set, but the object is not. It follows, then, that a major reason to prefer a scrambling structure over the non-scrambled one could be to allow the V as the focus, which it cannot be otherwise. As we saw, in English, the only way to obtain this result is to apply the marked stress rule, which shifts the stress to V. This was illustrated with (30), repeated in (39).

39 -Has your neighbor bought a desk already?
a) #-No, he is <u>building</u> a **desk**.
b) -No, he is **<u>building</u>** a desk.

[19]Zubizarreta (1994) surveys an unpublished paper of Truckenbrodt (1993), who found, essentially the same pattern in German, including contrasts like (35).

40 - Heeft je buurman al een buro gekocht?

a) #No, hij heeft in de tussentijd een **buro** getimerd

b) No, hij heeft een buro in de tussentijd **getimerd**
(No, he has a desk in the meanwhile built)

In Dutch, in the same context, there is no need to use the marked rule, and the scrambled version does the same job, as we see in (40).

With this, we may return to the contrastiveness generalization observed by de Hoop (1992), repeated:

9 A descriptive generalization:
"In Dutch, scrambling of the object yields the same semantic effect" as the contrastive predicates with stressed verbs in English. (dH, p. 165)

This descriptive generalization now follows from the analysis of stress and focus, and it is what we have just seen: Scrambling in Dutch, does the same job that the marked focus rule does in English. (We may note that the term 'contrastive' here may be misleading. What is true of the English equivalent is that it is obtained by the marked stress rule. This does not necessarily yield a contrastive reading, in the narrow sense, but allows the V to be a focus. I will return to this point in section 4.)

Let us see, next, how the major facts of object scrambling follow. First, there is a clear case where we expect scrambling not to be possible, namely, when the verb is an unlikely candidate to serve as the focus. Recall that of the cases de Hoop discussed. the clearest judgments were found precisely in these contexts, as e.g. in (6)- (7), repeated in (41).

41 a) omdat ik altijd een kat heb
because I a cat always have

b) *omdat ik een kat altijd heb (dH 72, p. 163)

c) omdat ik een kat altijd liefheb (dH 71, p. 163)
because I a cat always love

The verb <u>have</u> is a light verb that will require a very special context to serve as focus. The scrambling, in (41b), puts the main stress on it, thus making it the focus, which is why it is weird with no such special context. A verb like <u>love</u> is a content verb which can be focused, hence (41c) is fine.

Next, we may note that there exists another reason to use the marked stress rule in English. When the object is a pronoun, which cannot carry the stress, as in (42), the stress moves to the verb.

42 a) *Lucie saw **it**
b) Lucie **saw** it

Precisely the same is obtained in Dutch by use of the scrambled version, in which the main stress falls on the verb with no appeal to the marked stress rule.

43 a) *Ik heb gisteren het **gelezen**
b) Ik heb het gisteren **gelezen**

(43) cannot be explained by the definiteness condition (repeated in (44)). This condition only allows strong NPs to scramble, but does not require them to do so. (An arbitrary example with an unscrambled definite NP can be found in (36).)

The remaining question is what is the source of the definiteness effects that Dutch speakers associate with scrambling. Specifically, do we still need, in addition to what we derive so far, also something like the de Hoop's strength condition (5), repeated in (44)?

44) de Hoop's generalization: Only strong NPs can 'scramble'.

Recall that the intuition behind (44) is that, while definite NPs can always scramble (or not), in the case of indefinites, scrambling can be found in two cases: Either the indefinite is generic, or it is 'specific' (referential or partitive, in de Hoop's terms). The question of generics is, actually, independent. The genericity examples usually involve the operator always. It has been observed that scrambling, in this case, often results in truth conditional differences. (This was discussed extensively by Diesing and Kratzer, for German) Neeleman (1994) argues that in this case, there is an actual scope ambiguity, due to the different configurational relations of the operator and the object in the two structures. So the choice between the scrambled and the non scrambled option is motivated by semantic considerations of the intended scope construal. Let us, therefore, leave these aside here. The real issue is 'specificity', namely speakers feeling that a scrambled indefinite must be d-linked, previously mentioned, or otherwise more referential in nature. As we saw, it was feelings of this kind that led to an enormous enrichment of the syntactic system in order to capture them, in frameworks like de Hoop's or Diesing's.

In fact, if scrambling is motivated by focus considerations, this result is to be expected. Recall that scrambling, just like stressing of the verb in English, takes place if the verb has to become the focus, (or the object needs to be destressed). When would it be, normally, appropriate to use the verb alone as focus? The most standard situation is when the object has already been in the context set (or in the set of alternatives propositions in Rooth's (1992) analysis). Though this is a general conceptual consequence, it can be illustrated, for concreteness, with example (40), repeated. This is a typical context where it is appropriate to either stress the verb or scramble, and the object is, indeed, previously mentioned.

40 - Heeft je buurman al een buro gekocht?
(Has your neighbor already a desk bought?)

a) #No, hij heeft in de tussentijd een **buro** getimerd
b) No, hij heeft een buro in de tussentijd **getimerd**
(No, he has a desk in the meanwhile built)

In other cases the object itself may not have been mentioned, but its set was either mentioned or is inferable - what Pesetsky called d-linking. (E.g. in Ladd's example that I will return to: Has John read 'war and Peace'? No, John does not **read** books.) On the other hand, in the non scrambled structure where the object gets the main stress, it can serve as the focus. It is generally the case that indefinites are most easily used as foci, because in their most standard use they introduce new entities to the discourse (or represent 'new information'). So, the more standard use of indefinites remains the unscrambled version, where they are indeed the focus. It is easier and more common to find non-focus objects when they are definite, then indefinite. Hence, the greater ease for definites to occur in scrambling structures.

So, the intuitions regarding the specificity or definiteness effects of scrambling are, of course, correct, but they do not require a special condition to account for them. Rather, they follow from the focus analysis of scrambling.

As further evidence that the focus properties of the structures, rather than the definiteness factor, determine their acceptability, we may focus attention on the scrambling of definite NPs. This will show that, in fact, they are just as restricted as indefinite NPs. Scrambling is permitted only when the resulting focus structure is appropriate to the context.

45) Ik heb gisteren het **boek** gelezen
a. *en niet **verscheurd**
b. en niet de **krant**
I have yesterday the book read a. *and not tear up, b. and not the news paper.

46) Ik heb het boek gisteren **gelezen**
a. en niet **verscheurd**
b. ?en niet de **krant**
(same as (45), with scrambling.)

47 a) Dat Jan niet alleen zijn **leraren** prijst, maar ook de **decaan**
that John not only his teachers praises, but also the dean
b) *Dat Jan zijn leraren niet alleen **prijst**, maar ook de **dekaan**

Elliptic conjunctions, as in (45)-(47), signal the focus element, in the first conjunct, which must be the correlate of the second conjunct. In (45) - the non-scrambled version - neutral

stress marks the object as the focus. Hence the conjunction with the verb, in (45a), is out. In the scrambled version, (46), the situation is reversed. Here the verb is the focus. Hence, the conjunction with the verb, in (46a), is fine. The same is illustrated, with even a sharper contrast in (47). (All sentences should be read with neutral stress intonation.)

In (48), the structures are, appropriately, parallel, as far as stress goes. However, the comparison marks the underlined NPs as the foci. The neutral stress with scrambling falls on the verb. So we have a mismatch between stress and focus, and the sentence is hardly processable. In this context, only the unscrambled version in (49), can be used, where neutral stress falls on the NP which is indeed the focus.

48) *Ik heb de krant nog niet **gelezen**, maar ik heb het boek al wel **gelezen** .
I have the newspaper not yet read but I have the book already indeed read

49) Ik heb nog niet de **krant** gelezen, maar ik heb al wel het **boek** gelezen

As before, the judgments on focus appropriateness with scrambling are much sharper than any of the judgments of definite-effects that are cited in the scrambling literature. This, again, is consistent with the claim that scrambling is motivated by focus considerations, while the semantic (definiteness) properties of the scrambled NPs are only a (pale) reflection of the process.

4. The concept of markedness: Focus and economy.

4.1. Cinque's view of focus is striking in its simplicity and elegance. If it can be maintained, then focus is, essentially, a PF issue. Independent considerations of the computational system determine that stress must be assigned to a sentence. At the interface, this property of sentences is used to facilitate communication, using stress as focus. As we saw, this was, essentially, the view of focus in Chomsky (1971).

We should note, however, that the analysis is based on a revival of the concept of markedness, i.e. the idea that a distinction can be drawn between the neutral procedure of sentence stress, and other procedures which are marked. This distinction has been challenged extensively. It was repeatedly argued against the nuclear stress rule, or Chomsky's (1971) focus analysis, that in the appropriate context, main stress can fall anywhere, with effects hardly distinguishable from that of the neutral stress. This was particularly emphasized by Selkirk (1984), and I will return to the details of the arguments directly.

The crucial problem here is the same as that we observed in the case of QR and quantifier scope, namely, whether any content can be given to the concept of markedness. If it is just as

easy to construct examples with 'marked' stress, as with neutral one, and there is no obvious way to distinguish them, we run into the danger of vacuity - having a theory which excludes nothing: The facts that follow from its rules are labelled 'neutral', and everythign else - 'marked'. (This type of theory is always true, regardless of what its rules are.) A more realistic conclusion appears to be that there is no sentence-level generalization governing the selection of possible foci, and any expression can be a focus, subject only to discourse appropriateness. This, in fact, seems to have been the winning hypothesis for years, until Cinque reopened the issue. Possibly, this is also the reason why Chomsky (1976) departed from his earlier view, and took the position that focus-scope is determined just by QR. Any constituent permitted to raise by QR can, thus, serve as focus.

However, we noted already in the previous discussion of markedness, that it is a mistake to search the evidence for it in the realm of direct intuitions. A marked derivation is a derivation violating economy. When this is done with no reason, the result is visibly awkward. But if using the uneconomical derivation is, decisively, the only way to satisfy a certain interface need, the result sounds perfectly fine, and it is only indirectly that we can see that it is nevertheless marked, or uneconomical.

In the case of focus, we have already, at our disposal, some way to test the markedness hypothesis, when we look across languages. One of the findings of Cinque, and, mainly, Zubizarreta, is that if a language has the means to get a certain focus structure without applying the marked stress rule (say, by choosing an alternative permissible derivation), then its application yields visibly bad results in that language. The parallel case in a language like English, with very limited word-order options, may sound perfectly fine, with no visible evidence for markedness. One example was mentioned already in (32) and (33), repeated.

32 a) Johnson **died**
b) **Johnson** died

33 a) Johnson e' **morto**
b) E' morto **Johnson**
c) #**Johnson** e' morto.

As observed by Cinque, the Italian (33) sounds incomparably more awkward than its English counterpart (32b), although in both the marked stress rule has equally applied. This is so, since in Italian, the same focus needs could be satisfied with the structure (33b), with no application of the uneconomical operation, but in English, there is no other way to turn the subject into focus.

We may note now that the same is true also for the structures we examined in Dutch. As we will see in section 4.3, it is very easy to imagine a variety of contexts that permit the (uneconomical) stressing of the verb in the English (50a). If the same procedure relocates the stress from the object to the verb in the Dutch (50b), the result is highly marked. It is only

possible in a contrastive context (#, otherwise)[20]. This is so, since the same stress pattern can be obtained with scrambling, as in (50c), using neutral, rather than the marked stress.

50 a) I haven't **read** the paper yet. (marked stress)
b) (#)Ik heb nog niet de krant **gelezen.** (marked stress)
c) Ik heb de krant nog niet **gelezen.** (neutral stress)

51 a) I have **seen** him yesterday. (marked stress)
b) #Ik heb gisteren hem **gezien**. (marked stress)
c) Ik heb hem gisteren **gezien** (neutral stress)

Similarly, while destressing of the pronoun in the English (51) is completely natural, the same is hardly possible in Dutch, as in (51b). Scrambling should be used instead, as in (51c).

The only exception to this generalization, which I know, is the case of semantically empty objects like <u>someone</u> <u>anything</u> etc. A point I return to directly is that in English, these usually require shifting of the stress to the verb, as in (53). This is the one area where the same is true of Dutch - stressing of the verb rather than scrambling takes place. The same was observed by Hubert Haider to be true in German.

52 a) Have you **eaten** anything already?
Heb je al iets **gegeten**?
b) Have you **seen** anybody here?
Heb je hier iemand **gezien**?

53) (#)Heb je hier een vrouw **gezien**?
(Have you here a lady seen?)

This is not the general case with indefinites: (53), with just an indefinite object, is, as expected, marked, and it only allows a special contrastive use, just as (50b). Although I have no real account for (52), we should note that this situation arises in cases the object has no semantic role - it only marks the existence of a theta role, but does not supply any further information (The sentence will have precisely the same interpretation with the object omitted (<u>Have you eaten?</u>).

In view of this type of language variation, it seems that the concept of markedness is indeed necessary. Still, it may be useful to attempt a better understanding of what it means: how focus and markedness are defined, and how the definitions can be tested in a language like

[20]I am using the term 'contrastive' focus as defined in Rooth (1992): This is the case where the set of alternatives contains only one memeber.

English. So, I will turn now to a more detailed examination of these concepts.

4.2. Let me, first, summarize, and expand a bit, what I think is the real insight in the Cinque-based approach.

The basic point is that to identify the focus, two things have to be consulted: First, Cinque's generalized stress rule, which is a pure spell-out procedure completely independent of any other properties than structure. This stress is what the computational system provides its products with, at the spell-out stage, regardless of what use speakers may want to make of these products. Next, a procedure which we labelled the 'focus rule' relates stress and syntactic structures to focus.

Although Cinque may not have stated it precisely in the same way, I would like to elaborate a bit on the picture, which, I think, underlies this line of analysis: At the interface, sentences must be fit to context and purpose of use. One of the means relating sentences to discourse is focus. The computational system should provide us with sufficient means to identify the focus. This need has been often addressed by syntacticians with the idea of encoding focus syntactically: either by movement (QR), or by attaching a focus feature to nodes in the syntax, or both: attaching a focus feature to allow movement (which, interestingly, is viewed by some as more minimal than doing just one of these two). While certainly possible, this does not take us very far in addressing any of the problems discussed here, since we still need to know first, what the restrictions are on possible focus selections, and next, which focus selection is appropriate for which discourse.[21] Pursuing, instead, the line proposed by Cinque, each derivation is associated not with an actual focus, but with a set of possible foci, namely, a set of constituents that can serve as the focus of the derivation in a given context. This set is determined by the computational system at the stage where both the syntactic tree and stress are visible, namely, the focus rule applies either to a PF structure, or to a pair <PF,LF>, of sound and configurational structure.[22]. The members of the focus set it constructs are all and

[21]Once these questions are answered, the line of encoding a Focus feature in the syntax is a possible implementation. This, e.g. is the specific implementation chosen by Zubizarreta, who states the focus rule as a restriction on nodes marked +F(ocus). To deal with the problems I address here, this is unneeded. I leave open here the question whether there are other reasons to assume that +/-F is a syntactically encoded feature, as Zubizarreta argues.

[22]To get more precise about this description, we need to know more about the product of spell-out (namely on the nature of PF, in the pair <LF,PF>): Recall that Cinque assumes that at least as far as stress is concerned, it can be determined directly on syntactic structures, with no need to construct additional phonological structure. This is clearly the most minimal approach, and thus, the starting hypothesis that we would like to maintain, unless confronted with massive empirical evidence to the contrary. Still, this leaves us with two possible views of what PF is - one that this is just a sound string, the product of all spell out procedures. The other is that just like LF, this is the full syntactic

only syntactic subtrees (constituents) which contain the main stress of the derivation. It may, in fact, be more appropriate to view the focus rule as a definition of a set, rather than a procedure assigning features.

At the interface, one member of the focus set is selected, as the actual focus of the sentence. At this stage, it is up to discourse conditions, rather than syntax, to determine whether a derivation with a given focus is appropriate for a given context. If no member of the focus set can be used as focus in the given context, this derivation is unusable in that context.[23]

Summarizing, what gets uniquely determined by the computational system is the stress of the derivation, which is determined rigidly by the stress rule independently of either focus or context. The focus-set is then computed from the structure and the stress. Any stress pattern other than that determined by the stress rule is considered marked, and the next question is, then, the analysis of markedness.

4.3. As I mentioned, the concept of neutral sentence-stress or focus has been a subject of an extensive debate, fully acknowledged by both Cinque and Zubizarreta.. (For references, see Cinque, footnote 24, and, mainly, Selkirk (1984).) It was repeatedly argued that in the appropriate context, main stress can fall anywhere, and it is impossible to identify some patterns as more neutral than others. Let us view some of the arguments, focusing on the cases where stress falls, in English, on the verb rather than its complement.

A major context allowing that is argued to be when the complement is 'old information' (or D-linked, in current parlance), as in (54).

54 a) -Has John read 'Slaughterhouse five'?
-No, John doesn't **read** books. (Ladd, 1980, cited by Zubizarreta).

b) I'd give the money to Mary, but I don't **trust** mary. (Schmerling, 1976, cited by

---

tree, derived up to the stage of spell out, representing also further steps in the derivation required by spell out operations like stress, erasure of features, and other phonological processes. If the second is the correct view, then we may say that the focus rule applies solely at PF, namely, it associates a set of possible foci with each PF.

[23]This is consistent with the general view of the relations between syntax and use I proposed in Reinhart (1981). I argued there that each sentence is associated with a set of possible pragmatic assertions (PPA-set). The set is determined within the syntax, but discourse selection procedures determine which of these options, if any, is appropriate to a given context. I proposed there an algorithm only for determining the set of possible topic-predicate relations, but obviously, the full set of PPAs should contain also the possible foci of a derivation.

Cinque)

c) I can't **hear** you/him

In such cases, it may still be argued that the stress is not 'neutral' by the 'out of the blue' criterion. Such stress pattern requires a previous context - namely some mention of the complement or members of its set. But this is clearly not always the case. The verb gets the stress also with the most non D-linked of indefinite Nps, as in the sentences of (55), which can most typically be uttered 'out-of-the-blue'.

55 a) Are you **waiting** for anyone?
b) You should **eat** something.

Although I am not aware of any discussion of cases like (55), they
appear related to the contrast Bolinger (1972) found between the (a) and the (b) cases below (quoted by Zubizarreta and Cinque). In the (a) cases, the neutral candidate for stress does not merit a focus-status because it is semantically 'light' or un
informative. Here too, the sentences are perfectly utterable 'out of the blue' .

56 a) I have a **point** to make.
b) I have a point to **emphasize.**

57 a) I am going to the **doctor's** place.
b) I am going to the doctor's **barn**.

Cinque suggests another criterion for the distinction between 'sentence-grammar' main stress and 'discourse-grammar' (marked) stress: The latter allows only the actual category carrying it to serve as focus. E.g. by contrast to (27), of section 2, repeated in (58a). (58b), uttered with the marked stress on **red**, has only this adjective as its focus. (So, the only thing being questioned is the color of the shirt of the passenger the addressee is assumed to be looking for.)

58 a) Are you [looking for [a passenger with [a red [**shirt**]]]]?
b) Are you [looking for [a passenger with [a **red** [shirt]]]]?

In Zubizarreta's terms, the focus in such cases has only narrow scope (as is typically the case with contrastive focus), unlike the potential wide scope of the neutral focus. Indeed that is what my previous discussion in section 2 presupposed: The only option not included already in the neutral focus-set, as defined so far, is the one with the marked stress as the whole focus. While this may be true for this example, Selkirk and others have argued that it is not a necessary requirement in English. The arguments are surveyed in detail by Zubizarreta[24].

[24]Zubizarreta argues that the 'broader scope' option existing in English in such cases is still impossible in Romance. To account for such language- variations, she proposes (i)

Even in the case of NP modifiers, which (as noted by Zubizarreta,) are the hardest to use as wide-scope focus, we can find minimal pairs like (59).

59) a) If you are hungry, eat half a **loaf**.
b) If you are on a diet, eat **half** a loaf.

It would be hard to argue that any of the sentences in (59) (uttered independently of each other, and with no particular context) is more 'out of the blue', marked, contrastive, or narrow scope than the other[25].

To further illustrate this point, we may look at the sentence (60), based on an observation of Vallduvi (1990). (60) can certainly be used with the verb as a contrastive focus, if uttered, e.g. as reaction to the sentence Most people love spinach. But it can also be used in a context like (61). In this context it is clear that everything in the sentence, except for spinach, is included in the assertion, or the focus.

60) Most people **hate** spinach.

61) -We are having a **party** tomorrow. I think I'll cook **spinach**.
-Most people **hate** spinach.

If it is just as easy to construct examples with 'marked' stress, as with neutral one, and there is no obvious way to distinguish them, it is easy to see why the winning conclusion has been, as I mentioned, that there is no sentence-level generalization governing the selection of possible foci, and it is governed only by discourse appropriateness.

4.4. Nevertheless, I think that Cinque's position is correct, and his strikingly elegant analysis of focus can be maintained. Let us try, therefore, to explore further the relations between

---

i) Language specific PF constraint
A constituent marked F may not dominate deaccented material.

In the spinach example below, where the whole IP (minus the spinach) seems to be the focus, this constraint is violated, since the focus constituent dominates the object, which normally should be stressed and is deaccented here. But since the language specific (i) does not hold for English, this is allowed.

[25]The intuition that seems to me closest to explain the focus structure of the (b) case, is that the loaf does not play any significant role in the advice. The proposal is to eat half of anything one otherwise would eat a full portion of. I.e. Although the NP is not given, and it is not, independently, semantically empty or light, in this context it becomes close to semantic emptiness.

stress and focus which should enable an account also for the facts just discussed.

Cinque assumes that special (marked) stressed belongs to the territory of the procedures governing focus selection -His discourse-grammar rule (29) of section 2, creates a new focus (i.e. a focus not included in the standard focus-set, by changing the stress pattern. Presumably, this is a specific operation designed to turn the newly stressed unit into a focus. Hence we obtain the narrow scope, or locality, which Cinque associates with this procedure, which, as we saw, is the problematic consequence, at least for English. But it is not inherently necessary for the analysis to view marked focus this way. Note that what is visibly different with marked focus is the stress: The stress' position in fact violates the stress rule. We may assume that this is what is responsible for the marked nature of the derivation. (Cinque, noting that there is more than one function of marked stress, considers the idea that two distinct procedures may apply. I will suggest that the various functions can still be derived from one and the same operation.) Let us assume, then, that marked stress is obtained by a procedure relocating the main stress. Further technical details are not important for the present discussion, so the discourse procedure (29) can be simply restated as the minimally different (62), an operation which applies during the cyclical assignment of stress.

(62) Relocate the stress of a given cycle.

As before, stress relocation is an unneeded and uneconomical operation. A derivation to which it applied is, therefore, what we referred to as 'marked'. (It could also be described as ill-formed, as far as PF and the computational system are concerned.)

We may turn now to the procedure computing the focus set. On the view just outlined, this procedure need not care how the main stress was assigned. It applies the same way whether it is the standard neutral stress, or it is assigned by the special procedure (62). (This much, then, is consistent with the intuitions underlying the claim that any constituent(s) can equally function as focus.) However, more attention is needed to the formulation of this procedure, independently even of the question of markedness. In fact, the elegant idea that the focus-set contains only full constituents is not fully maintainable. One of the difficulties for any theory of focus, is that there can be several, non continuous constituents in each sentence that can serve as focus. This is obvious if we look at multiple <u>wh</u> questions and their answers, as in (63), where the focus sequence is the subject, the two objects, but not the verb, or the even more complex case in (64). (Examples with non <u>wh</u> contexts are given later, in (81))

63 a) -<u>Who</u> gave <u>what</u> to **<u>whom</u>**?
b) -<u>Max</u> gave <u>a book</u> to **<u>Lucie</u>**

64) Who thought that the letter we sent to whom may upset whom?

There are several options for capturing this, but for here, let us take the least interesting, but descriptively adequate approach, that the definition of the focus-set is as stated in (65), which

replaces the focus rule (24).

65) Focus set
The focus-set of IP contains any sequence of constituents of IP, which includes the main stress of IP.

4.5. Economy. So far this looks like a very permissive system. The marked stress operation feeds the procedure computing the focus set, which, itself, is rather permissive. So it seems that anything whatsoever should be allowed to serve as focus. How does this differ from the alternative view, that there are no PF restrictions on clausal stress or on the set of possible foci, with no notion of markedness involved? In other words, the question is what restricts this system.

Basically, markedness has in this system the meaning that it was given in early stages of theoretical linguistics. A marked operation is one which may be considered as yielding an inappropriate derivation. In current terms, this is so, because it involves less economical steps than competing derivations. Next,I will assume the concept of interface economy based on Golan, Fox, Reinhart, which I surveyed in part II. The basic idea is that the reference set for computing global economy considers only derivations with the same interface output. Focus, by definition, is a relation between a structure <PF, LF> and context. Two derivations are focus-equivalent if they express precisely the same focus- relation to a given context C. If this is the case, the more economical derivation rules out the less economical one.

A more intuitive way to express this principle is that one is allowed to resort to a marked, uneconomical operation, only if this is the only way to reach an interface goal. A similar idea can be stated also in the framework of optimality theory, developed in work by Grimshaw and by Pesetsky.[26]

Let us see, now, how this can account for the cases we considered. Sticking, for ease of exposition, to the cases with a marked stress on V, the generalization observed regarding their distribution is that they can be used for two reasons (or in two types of contexts): a. When the verb alone is focus - the case we have focused on in the previous sections, or b. When it is inappropriate to let the object serve as focus, either because it is given (D-linked) , or because it is not informative enough ('light'). But except for the object, anything else can serve as part of the focus, under this use. So, what is excluded by both types of use is a situation where the

[26]In a sense, this line attempts to explore the same concept of markedness considered here. Within optimality theory, a derivation with marked stress would be viewed as just a violation of the neutral stress rule, with no reference to an uneconomical operation. But the focus rule should be stated in such a way that given the appropriate context, meeting the focus rule may outweigh meeting the stress rule.

object is included in the focus. In fact, these two uses were observed already by Cinque. He toyed with the idea that the two functions of marked stress may be governed by two distinct procedures: One assigns extra prominence to a constituent we want to focus. In our case, it gives the verb a more prominent stress than the object. The other is the destressing of the object, as happens, most notably, when the object is a pronoun. Ideally, rather than assuming two distinct operations, we should be able to show that the availability of both results is entailed by the same focus-computation system, which I believe is the case.

The first function of marked-stress above - to allow the verb alone to be focus, is the easier one to explain, and it is essentially what we assumed throughout the discussion: The focus set of the neutral stress, still excludes the option of the V alone: The V sequence in this case does not include the main stress. Hence, in a context where V needs to be the focus, the neutral and marked derivations are never equivalent, so they do not compete.

Let us turn now to the other use of the marked stress - defocusing the object. To explore this point, let us take the simple minded illustration of focus below, where a yes/no question constitutes the alternative set for the answer. Hence, all and only constituents in the answer which replace a parallel constituent in the question are viewed as focused.

66)	Did	you	decide to buy	a **hat**?
	-No,	I	decided	to buy <u>a **coat**</u>
	-No,	I	decided	<u>to sell a **coat**</u>
	-No,	I	<u>was forced</u>	to buy <u>a **coat**</u>
	-No,	<u>Max</u>	decided	to buy <u>a **coat**</u>
	-No,	<u>Max</u>	<u>was forced</u>	to buy <u>a **coat**</u>
	-No,	<u>Max</u>	decided	<u>to sell</u> <u>a **coat**</u>

In (66), <u>coat</u> is, then, a focused constituent. It also carries stress, by the neutral stress procedure. The focus-set rule as restated, allows any subtree that contains <u>coat</u> to serve as topic. Indeed, in addition to <u>coat</u>, any sequence of constituents dominated by IP can be selected to be focused, with precisely the same stress pattern, and some of the options are illustrated. (As throughout, stress is marked with bold-face, focus - with underlining.)

Now, suppose we apply the marked stress operation, and relocate the main stress on any of the other focus constituents in these answers, as in (67).

67)	Did	you	decide to buy	a **hat**?
	#-No	I	decided	<u>to **sell**</u> <u>a coat</u>
	#-No	I	<u>was **forced**</u>	to buy <u>a coat</u>
	#-No,	<u>**Max**</u>	decided	to buy <u>a coat</u>

As far as the focus-set rule is concerned, nothing is changed, since the constituent carrying the main stress is focused, as required, and the rule allows any subtree dominating this stress to

serve as focus. This is precisely the reason why this is not allowed. To obtain the focus pattern in (67) we could have used, in each case, the alternative answer in (66), with neutral main stress. In the given context, there is, thus, no justification to use the marked derivation, so the neutral one wins.

68)	Did	you	decide to buy	a **hat**?	
	-No	I	decided	to **sell**	a hat.
	-No	I	was **forced**	to buy	a hat.
	-No	I	was forced	to **sell**	a hat.

Now, let us look at the answers in (68). Here the object hat is not focused. I.e. in this context, focus identification should select only one of the other, underlined, sequence of constituents. Had we left neutral stress on hat, the focus rule will not allow that. The focus is defined as containing the main stress. If this stress falls on hat, no constituent that does not include it will be identifiable as focus.

In sum, then, the focus-set created by applying the uneconomical stress procedure (62) intersects with the focus-set available by the neutral stress rule. The marked operation is allowed only if in a context C, a sequence not in the neutral focus set is the only one which can make the derivation usable in C. The one generalization in common to all such contexts is that the constituent on which neutral focus would have fallen needs to be excluded from the focus sequence[27].

---

[27]I should mention that there are still problems to be solved with the definition, and details of the focus set. The real generalization seems to be something like (i).

i) The possible focus is any sequence of the tree except what is c-commanded by the main stress.

ii) Did you decide to buy a **hat**?
-No I was **forced** to sell a hat.

Compare, e.g. 68, discussed in the text to (ii):
(ii) is bad, but my focus-set definition, and economy allow it.

One of the reasons I avoided defining the focus set as in (i) is that it would clearly exclude anything but a narrow- subject only interpretation in the subject stress cases like (iii)

iii a) **Truman** died
b) The **rent** is due
c) (What's this noise?)-The **water** is running.

But I should mention that it is not fully clear that the present formulation manages, indeed, to avoid this problem.

PART IV: TOPICS AND THE CONCEPTUAL INTERFACE.

1. Topics and presuppositions.

1.1. The problem: presuppositional effects inconsistent with logic.

Existential presuppositions are commonly assumed for definite descriptions and proper names. In most semantic frameworks it is assumed that this presupposition is built into the semantics of the determiner. Thus, Barwise and Cooper (1981) take this line for the definite determiners, such as the and both. E.g. the, as defined in (1), is viewed as a universal quantifier, with a restriction on the cardinality of the N set. If the N set is empty, as e.g. is the case with (2), the truth value is undefined, rather than true or false.

1) ||the||(N) = ||every||(N) iff |N| =1, undefined, otherwise.

2) The Dutch astronaut landed on Venus.

We should note that in this case, putting the restriction on the N-set is motivated independently of the issue of presupposition, by the cardinality restriction that it contains exactly one member (or exactly two, in the case of both)[28].

However, in recent years, a growing attention has been directed to the fact that other types of NPs may also carry existential presuppositions. This is most striking in the case of strong quantifiers. Let us first examine this case, and then turn, in the next section to presuppositional effects of indefinite NPs.

Given classical logic, (3a), as logically represented in (3b), should be true. (Since there are no American kings, the antecedent is false, hence the implication is true). The fact of the matter is, however, that this is not how speakers judge it. It looks undefined, with the air of a presuppositional failure.

3) a) Every American king lived in New York

b) (**A**x) (x is American king) ---> (x lived in NY).

Among the first to direct linguistic attention to this problem were De Jong and Verkuyl (1985). (Another line which I will mention again is Groenendijk and Stokhof's dissertation). De Jong and Verkuyl argued that since classical logic is inconsistent with speakers intuitions, this means that every in natural language is a different quantifier than in logic, carrying a

[28]Nevertheless, whether this type of presupposition should, indeed, be incorporated into the semantics, as Barwise and Cooper do, has been a topic of extensive debate. As we shall see, Strawson did not, in fact, take this line. But in view of the added cardinality restriction, the presuppositional approach has greater plausibility in this case, than the in the other cases we consider.

presupposition of the non emptiness of its common noun set. I.e., they propose to extend the line that Barwise and Cooper took for definite determiners to all strong determiners. The denotation they assume for every is, thus, that given in (4).

4) ||every||(N) = {X c E | X N = N}, N =/ 0, undefined otherwise.

Since in (3) the N set is empty, (4) entails that the sentence is undefined, rather than true, which is consistent with speakers judgments. This analysis has become, by now, almost standard in certain circles. (E.g. Diesing (1992) takes it for granted that all strong determiners are presuppositional, and proceeds to argue that weak determiners may be that too.)

Still, there are some problems with this analysis. While (4) assumes that natural language every has a different semantics than its classical logic analysis, there are contexts where it appears to be used precisely under its logical analysis. Most well known (and noted also by de Jong and Verkuyl) is 'definitional' or 'law-like' contexts, as (5).[29]

5 a) Every unicorn has exactly one horn.
b) All moving bodies not acted upon by external forces continue in a state of uniform motion in a straight line.
c) All trespassers on this land will be prosecuted.

In such contexts, it is irrelevant whether the N set has any members. Thus, (5a) is judged true, even though there are no unicorns. While definitional contexts may require some special account in any case[30], Lappin and Reinhart (1988) argue that the broader generalization is that every is used under its logical analysis whenever assessing whether the sentence is true or false does not involve an empirical assessment of the states of affairs in the (given) world. Another context they cite is tautologies or contradictions, as in (6).

6 a) Every unicorn is a unicorn.
b) Every unicorn is not a unicorn.

---

[29](5b and c) are quoted in Diesing (1992) p. 96, attributed to Strawson, Introduction to Logical Theory, London: Mouton, 1952. (5a) is from Lappin and Reinhart (1988).

[30]E.g. one needs to explain why (5a) is true, while (i) is false, on its definitional use.

i) Every unicorn has 3 legs.

By anything said here it is not obvious why (i) should not come out as true.

Speakers usually judge (6a) as true and (6b) as false, as follows from classical logic. If natural language every is different from the logical one, and has the presuppositonal analysis in (4), both should come out, equally, as undefined, and the judgments of (6) are a mystery.[31]

De Jong and Verkuyl's solution to this problem is to assume that there are, in fact, two entries for every in natural language. The presuppositoal (4) is the standard, more common one. In addition, the classical logic every exists as what they call a 'marked' option. Although they actually discuss only contexts such as (5a), they say that the 'marked' -logical every is used in non contingent context, which would also apply to (6).

As far as semantics is concerned, this line is not problematic. Adding a presupposition to a

---

[31]Diesing (1992, p.95-96) surveys a reply proposed by Kratzer (class-notes) to the arguments in Lappin and Reinhart against the presuppositional analysis of strong determiners. It addresses only the problem of 'law-like', or definitional universal quantifiers, in (5), and the idea is that their special status can be derived from their generic nature: The N-restriction functions as a restriction on the necessity operator assumed to be active in generic interpretations, under Kratzer's analysis. (This can be 'law of nature' necessity, or 'legal' necessity, etc.). A sentence like (5), then, gets (in Diesing's summary) the representation (i), which if I understand correctly Diesing's explanation, intends to mean that for any (relevant) world z, if this world contains unicorns, then every unicorn in z has exactly one horn. (So the presupposition is maintained in the standard way, for world z).

i) ([]:there are unicorns) (every unicorns has exactly one horn)

This is an interesting line. Let us assume that (once fully spelled out) it is correct, as far as semantics is concerned. There are still some questions to be asked: First, it is not clear how such representations could be derived compositionally. Specifically, how does the every restriction get to be bound by the necessity operator. Standard unselective binding, which is applicable to indefinites and bare plurals is not available here, since the determiner is strong, hence, the N variable is 'closed' and unreachable by an external operator. (Had we allowed unselective binding into the N-set of strong determiners, we would not have a way to block donkey anaphora in, e.g. *Every/some man who owns every donkey$_i$ feeds it$_i$.) Next, to the extent that such an interpretation is permitted, it should be available for all generic contexts alike. But non definitional universal generics seem to show the same 'presuppositional air' as other universals. E.g. though (ii) is generic, it is not saved from the air of a presupposition failure.

ii) All unicorns in the park tend to wake up early.

This can, perhaps, be handled if not all generics involve a necessity operator, as in Diesing's summary, but only the 'low-like' ones. But, in any case, this analysis does not extend to the contexts in (6). If the generalization is indeed that the presupposition impression is absent in non empirical contexts, this analysis cannot capture it.

strong quantifier, the result is still a strong quantifier. (This is explained briefly in section 2.) So the important distinction between strong and weak determiners is maintained, and the issue at stake here is only semantic judgments of truth and entailments. Nevertheless, we should note that introducing an ambiguity is, often, not a real solution to a problem. Rather it may turn out just a different way to state the problem. In the given case, the problem is that universal quantifiers do not always behave in natural language as predicted by classical logic. The logical interpretation is found in non contingent contexts, while in the others, they appear to be presuppositional. Concluding that there are two entries for each of these contexts does not take us any further than noting that such two contexts exist. The question still remains what is the relation between the two uses of strong quantifiers, and how do speakers decide which one to apply.

The question is sharpened once we observe that it is not only strong Nps that can give rise to a presuppositional impression (an observation which is, by now, widely recognized, due to work by Diesing, Enc, Pesetsky, and many others). Unlike the case of strong determiners, speakers vary dramatically in judgment of sentences like (7), given in isolation.

7 a) No American king lived in NY
b) Two American kings lived in NY.

Given the classical logical analysis of these sentences, (7a) should come out true, and (7b) false. But in informal checking of these sentences in various classes, only about half the participants came up with this judgment. The others judged both as undefined, as would follow from a presuppositional analysis. How do we account for those who get the presupposition effect here?

Unlike the case of strong determiners, adding a non-emptiness restriction (presupposition) to a weak determiner is not a trivial move, since it turns it into a strong determiner. (This is shown in detail in Lappin and Reinhart (1988), and will be summarized in footnote /9 (below (11).) One of the most fruitful distinctions ever discovered in the semantic research is that between weak and strong determiners, which has proved to be relevant for the computational system in many contexts (there sentences, extraposition, donkey anaphora, headless relatives, and many others)[32]. So, we clearly do not want to lose this distinction. Of course, there is always a way out, if we allow ambiguities all over the place, and, indeed, Diesing (1992) argues that all indefinites (cardinal) determiners are ambiguous between the standard-weak and the strong-presuppositional interpretation. Under her analysis, the fact that judgments vary on (7), would follow from these sentences being ambiguous. As we shall see, making this work, requires non trivial complications of the computational system. I will return to this issue in detail, but keeping in mind that ambiguity should always be our last resort, let us see, first, what an alternative account could look like.

[32]The role that the semantic properties of weak NPs play in deriving the distribution of 'extraposition'-structures and headless relatives is surveyed in Reinhart (1987), section 2).

1.2. Interface approaches.

Let us start with the problem of the apparent presuppositions of strong determiners, like every. The alternative approaches are based on the assumption that there is only one entry for every, --the standard universal quantifier. Its presuppositional effects should follow from its semantics at the interface. I.e. these effects are viewed as a problem of language use, rather than as a property encoded in the computational system (namely, encoded either in the syntax, or in the semantic definition of determiners). In the broader sense, this line follows Grice's position that when discrepancies are found between the logical meaning and natural language use, this does not necesssarily justify postulating different semantics for natural language expressions, but rather it may follow from pragmatic conditions on the way sentences are actually used. Still, there are several possible excecutions of this broad idea.

In fact, a similar line was taken, even for the case of definite-descriptions, by Strawson (1964). Strawson is often cited as the source of the idea that presuppositions are encoded in the semantics of the determiners. However, in this less widely cited paper, he pursues, in fact, a completely different position. He argues that a presupposition cannot be a property of a linguistic expression, but it is a property of the use of expressions in context - "as an when used in a statement with the role of identifying reference" (p. 94[33]). Crucially, the question is not what is the presupposition of the expression, but under what circumstances identifying reference fails, to the extent of giving the impression of a presupposition failure (- a "deficiency so radical as to deprive [the statement] from being either true or false", (p. 90), or a 'truth-value gap' in Quine's terminology, which Strawson endorses here). This, according to Strawson, does not happen every time an expression fails to refer (or has an empty N-set,in our terms). Rather, it is directly related to the actual procedure of assessment of a sentence uttered in a given context.

Strawson assumes (not unstandardly) that assessment is checking 'predication', where any expression in the sentence can be taken as the argument and the rest as predicate. Put in our terms, we may start assessment with any of the sets specified in the sentence. If assessment happens to start with checking an empty set, (or a definite description which fails to refer), we "gets stuck" -which is what a "presupposition failure" means[34]. The question is, of course, how the order of assessment (the set we start with) is determined.

---

[33]Page numbers for Strawson refer to the reprinted version in Sternberg and Jacobovits (1971).

[34]It should be obvious that I am not necessarily using here Strawson's own terminology. Strawson considers only definite descriptions, and coaches his analysis in the familiar terms of the truth-value gap theory: If we chose a non referring NP as the argument expression, there is no sense in asking whether the predicate, or its negation, apply to the denoted object, since there is no such object. Nevertheless, I believe that my description is consistent with the spirit of Strawson's analysis.

Strawson discusses in this paper only definite descriptions, and his answer is that the decisive factor is what the sentence is taken to be about, in the given context, namely, assessment always starts with the topic expression, where what the topic is, is determined by the context. If the king of France is bald is used to assert something about the king of France (with the king of France as topic), then to assess the sentence we have to search whether that king has or does not have the baldness property. But since we cannot find the entity to apply this to, we cannot continue. However, if the sentence has another topic, there is another assessment option. E.g., if we have been taking about a certain exhibition, (8) may not raise any problem of a 'truth-value gap'.

8) The exhibition was visited yesterday by the king of France.
(<2>, p. 95).

This is so, since assessment may start with checking our knowledge about the exhibition. Since we will not find any visit by the king of France there, we judge the sentence as false. (Technically, Strawson says that in such cases the failing NP is absorbed into the predicate.) In this case, it could appear that it is simply the availability of a successfully referring expression which enables assessment, but this is not so in the context (9).

9) A: What examples, if any, are there of famous contemporary figures who are bald?
B: The king of France is bald. (p. 96)

In such context, (9B) (with the king of France having a focus stress) is not an assertion about the king of France, but about the set of bald figures. Assessment will start with checking this set, and "since it is certainly false that [this] class ... includes any such item as our answer claims it does, the answer can, without too much squeamishness, be simply marked as wrong", i.e. the sentence is simply false.[35]

A problem with Strawsson's analysis is that it involves very subtle judgments, which were not, in fact, shared by many linguists. (E.g. they are challenged by Fodor (1979), and others in the same volume). It is extremely difficult to perceive any real difference in the presuppositional effects of the various contexts. The same is true also for the case of universal quantifiers that we are considering. If assessment is sensitive to topichood, we should get a contrast between (10a) and (10b): It should be easier to judge (10b) as true, than (10a). That this is not really the case is even easier to observe in these examples than in the case of definite descriptions (since the contrast involved is between a 'true' and 'untrue' judgment). The generalization seems to be as stated above - that this effect is found in all contexts involving an empirical

[35]That only topics carry presupposition was argued also in Gundel 1974, though along somewhat different lines.

assessment, regardless of whether the strong NP is used as topic.

10 a) Every American king visited Utrecht
b) Utrecht was visited by every American king.

In Lappin and Reinhart (1988), we attempted an analysis of the presuppositional effects of strong determiners, which keeps the essence of Strawson's approach, but independently of topichood.[36] We suggested that Strawsson is right that we get a presupposition failure iff the empirical assessment of the sentence must go through an empty set. But whether checking the N set of a given NP can or cannot be avoided (in the procedure of assessment) follows from the semantic definition of the determiners. It is only when the semantics permits assessment without checking the N-set that topic considerations may play a role. (I will argue directly that Strawson's topic analysis is relevant for the problem of presuppositions with weak NPs). Since the analysis is available, I will not repeat the technical details, or attempt to be fully precise here, but rather summarize the intuition behind it.

Our starting point in the assessment, following Strawson, are all the sets denoted by expressions in the sentence. Suppose one of them is empty, the question is whether the truth value of the sentence can still be determined without discovering that fact, or having to know anything about the cardinality of that set. The crucial semantic difference between strong and weak determiners is that to determine the truth value of a sentence with a weak NP argument, it is not necessary to know the cardinality of the N-set. Though this is well known, it may be useful to review it here.

Following Keenan (1987) (and leaving many aside, for the time being), we assume that a determiner is weak iff it is existential, symmetrical and intersective, strong otherwise. (Keenan introduces existentiality, illustrated in (11i) as the defining property, but Barwise and Cooper prove in their appendix that the three properties are equivalent.) The properties are illustrated (rather than defined) below:

11)
Existentiality (Keenan)
i) a. Two cats yawn <---> two cats that yawn exist.
b. Every cat yawns ~<---> Every cat that yawns exists.

Symmetry
ii) a. Two cats yawn <--> Two yawners are cats
b. Every cat yawns ~<---> Every yawner is a cat.

[36]This was proposed only for the non definite strong determiners. For the definite determiners, we assumed that the restriction on the N-set is built into the semantics, as in Barwise and Cooper. This is independently needed, for capturing the cardinality restriction on this set (uniqueness).

Intersection

iii) a. Two cats yawn <---> two cats that yawn yawn.

b. Every cat yawns ~<--> Every cat that yawns yawns.

The (a) pairs, with a weak NP are equivalent. But the (b) pairs, with a strong one are not. (In (ib), if it is true that every cat that yawns exists, it is not entailed that every cat yawns. Similarly, in (iiib), that every cat that yawns yawns, does not entail that every cat yawns.)[37]

The symmetry property of weak determiners is crucial for the assessment account: it entails that with weak NPs, we can always start assessment with the predicate set.

7) Two American kings lived in NY

To assess (7), repeated, we may either check the set of American kings and see if it contains two NY residents, or, by symmetry, we may check the set of NY residents, to see if two of them are American kings. Given Strawson's analysis, if we choose the first line, assessment gets stalled, and we get the impression of a presupposition failure. But if we choose the second, the sentence will be judged as false, (since the predicate American king does not apply to any of its members). Indeed, as we saw, both judgments were found in informal checking of this sentence.

But strong determiners are not symmetrical. Crucially, to assess a predication involving a strong NP, it is not sufficient to know the cardinality of the predicate set or of the intersection of the N and the predicate sets, but we need to know the cardinality of the N-set. To know whether most, or all, cats yawned is true, it is not sufficient to know how many cats yawned, but we need to know how many cats there are.

---

[37]Given symmetry, it is easy to show that adding a restriction on the N-set (presupposition) of weak determiners turns them into strong ones. If we do that, then D(A) is defined iff A is not empty. Let A be non-empty, and B be empty. Then || B c D(A) || is defined, but ||A c D(B)|| is not. Hence they are not equivalent. Hence, D is not symmetrical. Since the three properties above are equivalent, D is not existential, not intersective. I.e. it is strong.

E.g. if two is pressupositional, i.e. undefined when its N set is empty, the sentences in (i) below are not equivalent: (ia) is false, but (ib) is undefined.

i a) Two redheaded are American kings.

b) Two American kings are redheaded.

Hence, two is strong, just like every in (11).

Assuming, as we have, that in (3), repeated in (12), the sets relevant for assessment are the N set of American kings and the VP set of NY residents, it follows that it is impossible to assess the sentence without accessing the set of American kings. But this set is empty, so assessment cannot proceed.

12) Every American king lived in NY.

Since this is a semantic property of strong determiners, it does not matter whether another (topic) expression is available for assessment. In any case, the assessment must also go through the empty set.

If this analysis can be maintained, we answer all the problems posed so far: Presupposition effects always show up with strong determiners in empirical contexts, since their semantics requires that their common noun set be checked. This happens only in empirical contexts, since it is only there that the sentence is subject to the procedure of assessment (checking knowledge of the given world). With weak NPs, by contrast, assessment is not constrained by the semantics, hence we get the presupposition effect only if the option of checking the N set was selected.

Although I believe that this analysis gives the right results for weak determiners - a point I will return to - there are still some loose ends regarding the way it applies to the very case of strong determiners, for which it was proposed.[38] I should mention, therefore, that an alternative line

---

[38]A problem for Lappin and Reinhart's (1988) analysis is that if the procedure of assesment is taken to be the standard logical 'verification', there is always a way to assess a sentence with a universal quantifier by checking the complement set of the predicate. E.g. we can assess (i) by checking whether the set of non-NY residents contains any American kings, which will turn the sentence true, with no apparent presupposition failure.

i) All American kings were residents of New York.
ii) All American kings were not residents of New York.

Our answer was that economy considerations determine selecting the given predicate set rather than its complement set, since the former is usually smaller, hence requires less processing effort. This answer was mistaken: With negative predicates, as in (ii), it is often the case that the complement set is smaller than the given set (-there are more people who are not NY residents than people who are). Still, there is no difference in the assessment or presuppositional air of (i) and (ii).

It was probably a mistake to assume that the concept of assessment here is identical to logical verification and that all members of the equivalence set are equally available to the processor. Strawson himself was careful to allow assessment to consider only sets which are directly expressed in the sentence, or have been "antecedently introduced" (p. 96). If this is so, then we can not resort to complement sets here regardless of their size.

exists for this problem (of strong determiners), which also views it as an interface problem, but in a different way.

Groenendijk and Stokhof (1984) propose to view the presuppositional effects under consideration as an instance of what Grice called generalized conversational implicatures. (The analysis is also summarized in Gamut (1991, chapter 6). The general pattern of such implicatures is that a strong proposition (a) exists, which entails a weaker one (b). Use of (b), rather than (a), then, implicates the negation of (a). This is exemplified in the famous cases of (13).

13 a) Alfred is a pianist and a linguist ---->
b) Alfred is a pianist or a linguist.
c) Implicature of (b): Alfred is not both.

14 a) Lucie has more than three cats --->
b) Lucie has three cats.
c) Implicature of (b): Lucie has exactly/ (no more than) three cats.

G&S' analysis of every rests on the fact that ~**E**x P(x) entails **A**x (P(x) --> Q(x)). E.g. (15a) entails (15b).

15 a) There are no American kings -->
b) All American kins live in NY.

Hence the general pattern of such implicatures is met (avoiding (15a) implicates its negation, namely that there are American kings). While in the better studied scalar implicatures, the implicature is sensitive to the maxim of mannar ((a) and (b) should be equal with respect to mannar), G&S propose that the maxim operative in our case is that of relevance. If the stronger proposition is relevant, this is the one that should be used, and avoiding it implicates its negation. Computing relevance is based on forming a question for which, in the given context, the utterance of the given proposition is an answer. In their example (16), the question is given.

What is involved here has been a subject of much discussion in the philosophical literature regarding the concept of aboutness. Under the standard logical definition, assumed, e.g. by Putnam (1958) and Goodman (1972), "what the sentence is about is independent of which of its various equivalent formulations is employed" (Putnam, 1958:125). But it does not seem realistic that this concept of aboutness is what speakers assume when actually processing a sentence. This is discussed in Reinhart (1981), where I propose to use a pragmatic definition, which follows that of Strawson.

16) Child: Can I have a cookie?
Mother: You can have every cookie in the box.

If the mother knows that there are no cookies in the box, then that information is the most relevant one. Using the weaker answer in (16) would be a violation of the relevance maxim. Hence the child (assuming that no maxim is violated) can infer that the negation of the stronger proposition is true, namely that the common noun set is not empty, which is what we called an existential presupposition.

Turning to to the problem I discussed here of why these presuppositions appear only in empirical contexts, we may assume that in non-empirical contexts, such as definitions, whether the N-set has any members is not relevant. E.g. if the child asks, seing a picture of a bull "Is this a unicorn?", and the mother answers, "No, all unicorns have just one horn", the existence of unicorns is not relevant to the conversation, hence nothing is inferred about it from the mother's utterance.

Obviously, an account resting on the concept of relevance is bound to leave many questions open as well. I will leave these questions untouched, since my major concern here is weak-existential NPs. At worst, we may conclude that the presuppositonal effects of strong determiners are still a problem. However, the problem remains preciesly the same if we encode it in the computational system and declare strong determiners ambiguous. Whether they are ambiguous or not, the question in which contexts they are used with presuppositional effects and in which not, is still an interface problem.

1.3. Topics and the assesment of existential NPs.

The area where a Strawson based approach seems most profitable and unproblematic is that of weak determiners. It is a solid conclusion that the semantics of weak determiners does not restrict their assessment in any way, so further restrictions may be posed by discourse considerations.

So far we considered sentences like (7), repeated in (17), and assumed that given no context, speakers chose arbitrarily how to assess them. Concequently, about half the class-participants the sentences were checked with judged (a) as true and (b) as false; the others judged both as undefined.

17 a) No American king lived in NY
b) Two American kings lived in NY.

18 a) There were no American kings in NY
b) There were two American kings in NY

But, interestngly, the situation is quite different in (18). Here there was no variation in judgments, and all participants judged (18a) as true and (18b) as false, in accordance with their logical analysis. How can this be explained within the interface (assessment) approach?

Recall that Strawson's original account assumed that the order of assessment is determined by topic considerations: Assessment always starts with the expression that is viewed as the topic in the given discourse. In the case of strong determiners, we saw that the judgments did not conform with this analysis, and there must be an independent factor that determines that the N-set cannot be skipped, when the sentence is subject to empirical assessment. However, no such factors force assessing the N-set with weak determiners. So, it is in this area that the predictions of Strawson can be checked. Even in view of the massive varieties of opinions regarding what topics are, (3) is the one context all studies agree upon: The NP in there sentences can never be topic. (These are the structures that the Prague functionalists, and others, like Kuno, called 'presentational'.)
So if assessment starts with the topic, it could never start with the existential in (18).

In (17), by contrast, there is nothing inherent in the sentence that can tell us whether the existential subject is topic or not. Here, the theoretical analysis of what topics are plays a greater role. The traditional studies of topic (like the Prague-school) assumed that the defining property of topics is familiarity (being 'old information'). If this view is right, then, by definition, the 'new information' subjects in (17) cannot be topics. But in Reinhart (1981) I argue in detail that the familiarity approach to topics is mistaken. It is, of course, true that the majority of expressions identified as topics in actual discourse represent a 'given' entity (i.e. an entity already in the context set), or a familiar one. But this does not mean much, since it is equally true of all expressions in actual discourse. (Typically, a sentence in actual, recorded or written, discourse refers to several given entities, only one of which is the topic.) So, familiarity cannot be a sufficient condition for topichood. Furthermore, based on both textual analysis and a survey of a large body of studies of topics, I argued that it is also not a necessary condition. It is true that the common way to introduce new entities into the discourse is in focus position, or by use of presentational sentences, but an existing alternative is introducing them as (indefinite) topics. This is particularly common at the begining of a new segment, or following a semantic connective. (This is argued and exemplified in detail in that paper, so I will not repeat the discussion here.[39])

---

[39]Among the ways to assess this claim is checking backward anaphora in actual discourse. Typically, this is possible only when the antecedent is a topic (as argued by Kuno). But contrary to the common arm-chair beliefs that linguists tend to impose on discourse, backwards anaphora is easily found in actual discourse with new entities, including indefinites. Many such examples are listed in the corpus of Carden (1982). E.g.

i) "Because they wanted to know more about the ocean's current, students in the science club at Mark Twain Junior High School of Coney Island gave ten bottles with return address card

The difference between (17) and (18), then, is that in (17) the subject can be a topic, in an appropriate context, though it does not have to be. But in (18), it cannot be a topic in any context. Given that the sentences were checked in isolation, the decision whether to view the subject of (17) as topic or not, is arbitrary. Speakers who chose the topic interpretation, namely started assessment with this NP, came up with the undefined judgment.

This can be explained in a more precise way, given the full analysis of sentence topics in Reinhart (1981), which I will only sketch here. Each sentence is associated with a set of what I called there Possible Pragmatic Assertions (the PPA set). The members of this set are, first the bare proposition (in case there is no topic), and then, all possible pairs of an entity and the proposition (similarly to the cards of DRT). The set is restricted by sentence level considerations: While a normal SVO sentence has three members in its PPA set (SVO, S/SVO, O/SVO), there are structures where only some of the construals are possible. E.g. there sentences do not have a PPA with the subject as a topic. Passive sentences allow only (the bare proposition and) the subject as a topic. Left dislocation sentences mark the topic explicitly, hence have only one member, etc. In actual context, one member of the PPA set is selected, relative to that context[40].

The PPA sets of (17a) and (18a), are, then, illustrated, roughly, in (19)-(20)

19 PPA set of (17b):
{a. Two American kings lived in New York
b. Two American kings(x)\ x lived in New York
c. New York (y) \ Two American kings lived in y }

20 PPA set of (18b):
{a. There were two American kings in New York
b. New York (x) \ There were two American kings in x. }

In the absence of context, any member of the PPA set can be arbitrarily selected. There is only

---

insede to crewmen of the NYcity's sludge barges" (The New York Times)

ii) "When wshe was five years old, a child of my acquaintance announed a theory that she was acquainted by rabbits". (New York Times)

[40]The selection function, which is also defined in that paper, preferrs, e.g. a pair the topic of which is familiar, unless the sentence begins a new segment. If there are further choices, the subject is preferred over the object, etc. Obviously, the selection function, which is a real discourse procedure, is less easy to define than the PPA set. Nevertheless, I believe that a lot more can be said even about this procedure than is commonly assumed.

one selection which will lead to the impression of presupposition failure, namely (19b), where assessment will start with the empty set of American kings. It is, thus, only in the case of (17) that such an effect was noted, and only by an arbitrary number of speakers who selected this option.

When attempting to extend this analysis to other cases, we should note that this should be done with great care. It is extremely difficult to decide what are the facts in this area of presuppositins. In the particular examples that I discuss, the fact that there is a real issue of difference in truth intuitions was checked with a sample of speakers[41]. In most discussions of such effects (most notabely, by Diesing (1992)), the authors just declare their feelings regarding the presuppositionial status, and no way to actually verify these feelings is offered. Let me just exemplify here one context which was claimed to be relevant.

Diesing (1992) argues that subject scrambling in Gernan can distinguish the presuppositonal and non presuppositional interpretation of indefinites. I will return to the syntactic details of these structures in section 2, but the basic claim is that (21), which according to Diesing is ambiguous in English, can be disambiguated in German: (22a) with the VP subject is only "existential", while (22b), with an IP subject, is only presuppositional.

21) Two cellists have indeed taken rooms in this hotel.

22 a) ...weil ja doch zwei Cellisten in disem Hotel abgestiegen sind.
b) ...weil zwei Cellisten ja doch in disem Hotel abgestiegen sind.

Diesing does not provide any evidence for her feelings regarding these sentences. (Her description of her intuitions is surveyed in detail in a footnote in section 2.) But, given the method used here, her intuitions could be checked by replacing two cellists with an NP with an empty N set, like two/no American kings, and then subjecting them to some empirical testing with German speakers (which I have not done). If she is right, we should get a true/false judgment on (22a), and an undefined judgment on (22b).

If, indeed, there is a distinction here, it may correlate with the possibility that the 'scrambled' IP position of the subject in (22b) is a topic position, namely, that the motivation for choosing (22b), rather than (22a) is to signal the topic status of the subject. This too should be checked with great care, since there are only very few known structures where a positon is obligatorily topic or obligatorily non-topic. (In most cases, this is just a matter of preference). There sentences have been shown again and again not to allow their subjects to serve as topics. It is

[41]For the present problem, it is particularly important to do this rather than rely only on introspection. particularly since in the standard cases judgments vary. The difference between the two contexts here (17) and (18) could not hae been found by introspection, since for each individual speaker (17a) will not be perceived as ambiguous, and it would be judged as either true or undefined.

not necessarily true for other VP internal subjects. Although I did not discuss this here, there are tests to decide when a position is necessarily topical. Several tests are discussed in Reinhart (1981), but the most decisive test is provided by accessibility theory, which I discuss in section 3, below[42].

Finally, in case the assessment account sounds cumbersome, let me mention another way that could, perhaps, be developed to describe the intuition behind it. The generalization is that in the case of weak determiners, we get a presupposition impression iff the NP is construed as topic. NP topics are, intuitively, the entities we are talking about. It is a tendency of conversation not to indulge in discussions about empty sets. So we assume that in a felicitous discourse our topic entities exist, and we get confused when they turn out not to. We may, then, label this reaction 'an impression of a presupposition failure'.

[42]There are other options regarding what to expect of (22). It seems to me more realistic that the lower subject in (22a) can be either topic or not. Hence, it is similar, in this respect, to the English structure (21) and (17), rather than to there sentences. Hence we should get variation of judgments here. But it could still be the case that in (22b) there is only topic interpretation, so we could get a majority of 'undefined' feelings. Generally, I should add, it would not be realistic to find an undefined judgment unanimously in any structure, since knowledge of logic may interfere at least for some speakers.

2. Diesing's alternative of syntactic encoding of indefinite presuppositon.

My account for the apparent presuppositions of weak NPs following Strawson's insight, rests crucially on the notion of topic, which is a discourse notion. As Strawson showed, what the topic of the sentence is, in the sense relevant to our problem, can be determined only in context. Whether the analysis can be made fully explicit, then, depends on whether we could have an explicit theory of identifying the topic. Although I believe that the answer is far less vague and circular than standardly assumed in current syntactic literature, it is still the case that discourse notions are bound to never be as explicit and formal as the notions of a (good) syntactic theory can be. For this reason, many syntacticians (and semanticians) prefer to avoid working with discourse notions.

I believe that it is this worry about the vagueness of discourse theory that has been leading many syntacticians, recently, to attempt syntactic (or semantic) solutions to the 'specificity' problem of indefiniets, which was previously considered a discourse problem. Notable recent examples are de Hoop (1992), and Diesing (1992). Diesing addresses specifically the problem of presuppositions of weak determiners, and argues that this is, in fact, a syntactic problem, which is governed by a rich mechanism encoded into the computational system.

Under Diesing's analysis the examples discussed in this section should follow from what she calls 'the mapping hypothesis.' (The details of her analysis are provided directly). Briefly, existential presuppositions are possible in her analysis only when the N occurs in the restrictive term of an operator, and this is possible only if it is in an IP position. <u>There</u> sentences allow only for existential closure of the indefinites, since, more generally, the domain of existential closure is the VP[43].

It is important to note that what Diesing offers is a novel syntactic theory, and it should, therefore, be evaluated not only by how well it captures the discourse behavior of indefinites, but by its syntactic plausibility. A central claim of her analysis is that parametric variations between languages can stated at LF. Specifically, that while in English NPs can lower or raise at LF, in German, they essentially cannot. (That this is indeed what Diesing is assuming will be shown in the next section). This claim stands in sharp contrast to the most fundamental assumptions of current syntax, in which LF is taken to be universal. In the minimalist program, this assumption is the basis for accounting for word-order variations: The operations deriving LFs are universal, and languages differ only in where, during the derivation, phonetic

[43]The fact of the matter is that although the lack of presuppositional effects in <u>there</u> contexts is considered the strongest evidence for Diesing's analysis, she herself does not explain anywhere how this facts follows. E.g. what prevents the existential subject from moving, at LF, to IP, replacing the explitive? That this happens at LF is very commonly assumed in syntactic theory. Since Diesing assumes that in English the mapping hypothesis applies at LF, it is not obvious to me how this basic fact is derived. But let us assume here it is.

spell out applies. But even in previous stages of the theory, LF was never considered the level at which parametric variations can be stated.

For a syntactician, then, the question whether Diesing's account can be adopted, as stated, does not even arise. One does not abondon the foundations of the theory every time a new problem (like presuppositional effects in German) is discovered, unless it is shown that the alternative theory can account for everything the present theory can, and not just for the one new problem.

Nevertheless, this could be viewed as just a matter of implementation. The serious question brought up by Diesing's work is whether the presuppositional effects of indefinites should be syntactically encoded. If this is so, an analysis more consistent wiht current syntax could, perhaps, be sought. I would like, therefore, to examine here in detail the empirical aspects of Diesing's claim, abstracting away from an evaluation of her syntax. The question will be whether she provides any evidence that presuppositional effects correspond to syntactic ambiguity.

2.1. Diesing's mapping hypothesis.

2.1.1. Recall that we examined two approaches to the issue of presupposition. The one is encoding them into the semantics of the determiners. This line which was assumed by Barwise and Cooper only for the definite determiner, has been extended, e.g. by de Jong and Verkuyl (1985), to all strong determiners. The denotation they subsequently, assumed for every, is repeated in (1).

1) ||every N|| = {X c E | X N = N}, N =/ 0, undefined otherwise.

The alternative view is based on Strawson's (1964), for whom presupposition, to begin with, cannot be determined by the semantics of the determiner in isolation, but is a property of the use of Nps to refer, in context. Whenever assessment of a given sentence must start with checking the common Noun set of a given NP, if that N is empty, assessment "gets stuck" - which is what a "presupposition failure" means. Strawson assumes that what NP assessment starts with is determined by the context, and it is the NP linguists call 'sentence topic'. Whether this is true for all NPs, including definite and strong NPs, as Strawson argued, has been a subject of debate. However, we saw that this analysis is most suitable for weak (indefinite) NPs. The sentences which best illustrate this are repeated below. (In (2) truth judgments vary, since the subject can be taken as a topic, or not. But in (3), where it cannot be a topic, the truth judgment is as predited by logic.)

2 a) No American king lived in NY
b) Two American kings lived in NY.

3) a) There were no American kings in NY

b) There were two American kings in NY

An alternative approach, closer to the first view, is taken by Diesing (1992). She argues that presuppositions are encoded in the computational system. In the case of indefinites, they are all ambiguous between the standard weak reading and a strong, presuppositional, reading. Each of these corresponds to a specific syntactic position. In (2), syntax allows both structures, hence the sentences are ambiguous. But in (3), the structure associated with strong NPs is disallowed. Let us review her basic assumptions and machinary.

The LF of (4) is standardly assumed to be (4b), which is isomorphic to its representation in classical logic, in (4c). (An operation deriving 4c syntactically from (4b) by raising the determiner is proposed in Heim (1982).)

4 a) Every cat sneezed
b) [Every cat]$_i$ [e$_i$ sneezed]
c) (**A** x) (cat (x)) ---> (sneezed (x))

5 a) Some cat sneezed
b) [Some cat]$_i$ [e$_i$ sneezed]
c) (**E** x) (cat (x)) & (sneezed (x))

Following Heim, the first constituent in (4b) corresponds to the restrictive term, and the second -to the nuclear scope. Technically, the same can be said about the parallel structure with an indefinite NP in (5), except that here these terms do not mean much, given standard logic. Unlike the case of (4), the two clauses here are interpreted as a conjunction, and the order of the conjuncts could not possibly matter. (This is the symmetry property of weak determiners.) However, Diesing's point (following a growing consensus in DRT) is that this is not always so. In fact, a sentence like (5a) can receive two interpretations, corresponding to two distinct LFs. The standard conjunction interpretation (5c) is obtained only when the subject is VP-internal at LF. If an LF like (5b) is derived (or the subject occurs at SPEC IP at LF), the NP forms a restrictive term. The interpretation Diesing gives to this notion is that non emptiness of the N set is presupposed. Though Diesing does not formally represent the reading she intends, it appears that the intended reading can be represented in (6)[44]. (The indefinite is not bound here by any other operator, hence it must be closed by a discourse existential.)

6) (**E** x) (cat (x)) & (sneezed (x))
undefined if {cats} = 0.

The variation of judgment regarding (2a) will be explained in this approach by the fact that the sentence is ambiguous. In the presuppositional reading, (7a), it turns out undefined, while in

[44]See Diesing p. 62 for the semantics she is assuming for presuppositions.

the existential reading (7b) it is true.

7 a) (~**E** x) (x is an American king & x lived in NY), undefined if {American kings} =0

b) (~**E** x) (x is an American king & x lived in NY)

Whether the Strawsonian line or the line of syntactic encoding are correct for presuppositions, depends, of course, on how well the syntactic account is independently motivated. Let us examine, therefore some further details of Diesing's motivation.

2.2. The hypotheses Diesing is using are summarized below:[45]

8 a) The restrictive term is presuppositional: Non emptiness of the N set is presupposed iff it occurs in the restrictive term (of a tripartite tree, or formula).

b) The mapping hypothesis
Material from VP is mapped into the nuclear scope
Material from IP is mapped into a restrictive clause.

How do these assumptions generate the two readings Diesing assumes for the sentence we considered, repeated in (9)? As is standard, subjects (of stage -level predicates) are assumed to be base generated in SPEC VP, and surface, in English, in SPEC IP. From the latter position, they can lower, or reconstruct at LF back to their VP position. It follows that in the standard cases indefinite subjects always correspond to the two LFs illustrated in (9).

9 No American kings lived in NY
a) [^{IP} No American kings [^{VP} e lived in NY]
b) [^{IP} e [^{VP} No American kings lived in NY]

The mapping hypothesis (8b) now determines that in the structure (a) the indefinite is translated as the restrictive term. Combined with (8a), it must, then, be presuppositional, so it gets the interpretation (7a). In (9b), the mapping hypothesis determines that the indefinite is included in the nuclear scope. Hence, by (8a) it cannot be presuppositional, and it gets the

[45]While the mapping hypothesis (8b) is explicitly formulated by Diesing (e.g. p. 10), the correlation between restrictive terms and presupposition (which is the crucial point for the present discussion) is only stated informally. However, that Diesing intends something like the strong iff condition in (8a) is obvious from various points in the discussion, e.g. p. 62-23, 69.

standard existential interpretation in (7b). (It also follows from (8a) that NPs with strong determiners, like <u>every</u> can never lower back to the VP, since they are always presuppositional, hence they are interpretable only when attached to IP).

In the examples so far, the issue seems purely conceptual, and there is no way to assess the claim that the subject reconstruction correlates with any semantic distinction between the two structures. Diesing does provide massive empirical evidence for her analysis. However, a closer examination reveals that the evidence evolves around a different ambiguity than the one we have been considering so far -that between generic and existential readings of bare plurals, illustrated in (10)

10 a) Firemen are available
b) Firemen are brave.

(10a) has two interpretations. The existential one is that (at the given moment of utterance) there are firemen available (not necessarily at any other time). The generic interpretation is a statement about a defining property of firemen, namely that they are always available. (10b), on the other hand, has only one interpretation - the generic one. The semantic distinction here is clear and solid, and Diesing derives it as follows:

Kratzer (1989) has argued that the subjects of individual-level predicates must be base generated in SPEC IP (rather than SPEC VP, where subjects of stage level predicates are generated). Using an impressive body of evidence and argumentation, Diesing extends this to the claim that such subjects cannot lower to VP at LF. She argues that a distinction analogous to that between raising and control structures is needed here: In the case of individual-level subjects, the VP SPEC is occupied by a PRO, controlled by the IP subject. The SS of the two types is illustrated in (11b) and (12b).

11 a) Firemen are available
b) SS: $[_{IP}$ Firemen$_i$ $[_{VP}$ e_i are available]]

LF: c) $[_{IP}$ Firemen$_i$ $[_{VP}$ e_i are available]]
d) $[_{IP}$ e_i $[_{VP}$ firemen$_i$ are available]]

12 a) Firemen are brave.
b) SS: $[_{IP}$ Firemen$_i$ $[_{VP}$ PRO$_i$ are brave]]

While raising subjects can reconstruct back to the VP at LF, the subjects of control cannot do that. (This is based on arguments of May, regarding scope reconstruction into raising and control complements.) It follows, then, that in the case of stage-level predicates two LF derivations are permitted, as in (11c,d), while in the individual-level case, there is only one LF

derivation, identical to (12b).[46]

The semantic difference between the two structures follows in Diesing's framework, if we assume the analysis of the generic interpretation in the DRT framework (Kratzer, Heim), namely that bare plurals are standard indefinites, and the generic operator unselectively binds their variable. Implicit already in Heim (1982) is the condition (13)[47]

13) An operator unselectively binds a free variable, x, iff (an occurrence of) x is in its restrictive term.

Combined with Diesing's mapping hypothesis (8b), it follows that only when the subject is dominated by IP, a generic interpretation is possible. That's why the LFs (11c) and (12b) both allow the generic interpretation (14) (ignoring here the details of the time variable).

14 (**GEN** x) (x is a fireman) ---> (x is available/brave)

On this, then, bare plural subjects of stage and individual level predicates behave the same. However, stage- level predicates also allow the subject to be VP internal, as in (11b). This structure is interpreted by existential closure, yielding the existential reading in (15).

15) (**E** x) (x is a firemen & x is available).

Since the individual-level subject cannot occur VP initially, this interpretation cannot be derived for (12). Let us assume, then, that it follows that bare-plural subjects of individual level predicates allow only the generic interpretation.[48]

---

[46]The subject could also undergo further QR, however, this is clearly not necessary, since the subject occurs already in an IP position, and there is a relevant variable for it in SPEC VP. If there is a further QR operation, the interpretation will be identical, as far as the tripartite distinction is concerned.

[47] It is not fully clear that it is indeed necessary to assume in Heim's framework the strong version of this principle, as below. On this issue as well, Diesing does not actually state this principle explicitly and formally. However, it seems that she must be assuming at least that. Otherwise it is hard to see why the existential interpretation is not available in the restrictive term, nevertheless (i.e. why the individual -level sentences are not ambiguous as well). I should add that even under this formulation, this question is not fully clear to me - see the next footnote.

[48]In fact, it is not fully clear to me how the consequence that (12b) has only the generic reading follows. Given all the explicit assumptions above, it follows that the individual level subject in (12b) cannot be interpreted as a standard existential, as in (15). But Diesing's central point was, as we saw, that IP-dominated indefinites allow also the presuppositional reading which is not generic, as in the analysis of Two/some cats sneezed above. So, this structure could still permit the reading

Another evidence for the correlation between generic interpretation and the syntactic position in IP comes from German. Diesing argues that in German, subjects can surface either in IP or in VP -SPEC and their position can be witnessed. Assuming that sentential particle, like ja doch occur in the I projection (or "mark the VP boundary"), the subject in (16a) is in SPEC IP, while in (16b), it is in SPEC VP. (This is Diesing's example (25), p. 31.) Diesing shows further that extraction is possible from subjects following the particle (as in (16a), but not from those preceding it. This follows in her analysis from the first being in VP, hence L-marked by the V-I complex (at LF).

16 a) [$_{CP}$ weil [$_{IP}$ Ameisen ja doch [$_{VP}$ einen Postbeamten gevissen haben]]]
since ants 'indeed' a postman have bitten.

b) [$_{CP}$ weil [$_{IP}$ ja doch [$_{VP}$ Ameisen einen Postbeamten gevissen haben]]]
since 'indeed' ants a postman have bitten.

It follows, next, that subjects of individual level predicates can occur only to the left of the sentential particle, since in the post-particle position they would be in the VP, and given the Kratzer-Diesing generalization above, individual-level subjects must be in SPEC IP. This is shown to be correct in (17) (Diesing's 37a, p. 38). The mapping hypothesis (8b) suggests now, correctly, that in (17a), only the generic interpretation is possible.

---

in (i), which is the new presuppositional existential introduced in this work.

i) (**E** X) ({firemen} =/ 0 & x is a firemen) & (x is brave).

But it is clear that sentence (12) cannot have this reading, and it requires the generic interpretation. I imagine that to make the analysis go through, further assumptions are needed either on the nature of bare plurals, or some requirement that when a generic operator can apply it must do so.

That some such restriction must be implicitly assumed in Diesing's analysis is obvious also from her discussion (p. 81) of the Dutch case in (i), (from Reuland, 1988).

i) *Fred denkt dat koeien op het dak liggen.
Fred thinks that cows on the roof lie.

Diesing shows that the embedded subject here must be in IP at LF. Hence, "this leads to a generic interpretation of the subject. However this generic interpretation is pragmatically incompatible with the locative predicate 'on the roof', leading to the judgment of ungrammaticality" (p. 82). I.e. Diesing must be assuming without argument that only the generic reading is possible here, otherwise the sentence should have still be good on the 'presupposition of existence' reading.

17 a) [$_{CP}$ weil [$_{IP}$ Wildschweine ja doch [$_{VP}$ inteligent sind]]]
since wildboars 'indeed' are inteligent

b) *?[$_{CP}$ weil [$_{IP}$ ja doch [$_{VP}$ Wildschweine inteligent sind]]]
since 'indeed' wildboars are inteligent

Diesing argues, further, that German provides even a more direct evidence for the operation of the mapping hypothesis in the case of the generic interpretation. It turns out that in the IP-position, bare plurals always allow only the generic interpretation, regardless of the semantics of the predicate. E.g. in (18) (Diesing's (34a), p. 37), the stage-level predicate allows its subject either to occur in the VP, as in (18a), or to raise to IP. According to Diesing, in the first case it has only the existential reading, while in the second - only the generic reading.

18 a) [$_{CP}$ weil [$_{IP}$ ja doch [$_{VP}$ Kinder auf der Strasse spielen]]]
since 'indeed' children on the street play

b) [$_{CP}$ weil [$_{IP}$ Kinder ja doch [$_{VP}$ auf der Strasse spielen]]]
since children 'indeed' on the street play

The last observation does not yet follow from anything said (since the subject could further raise or lower at LF to provide the alternative reading for each structure). But Diesing adds (without much discussion) the stipulation that "unlike what happens in English, in German the tree splitting can occur at S-structure. In other words ... abstract movement operations such as LF lowering need not occur" (p. 39). If we change 'can occur' in this informal formulation to 'must occur', i.e. neither lowering nor raising are permitted at LF,in German (rather than 'need not occur'), the judgments of (18) are derived[49]. It might be useful, then, for the sake of explicitness, to add this parametric assumption to the theoretical machinery assumed:

8' c) <u>Parameters</u>: English allows LF movement, Dutch and German do not.

We may recall, in passim, that this parameter is strictly inconsistent with current syntactic theory, but we decided to ignore such matters here.

[49]The same 'parametric' variation is assumed also for Dutch, in the discussion of which Diesing is more explicit, stating that "indefinite subjects cannot undergo LF lowering in Dutch" (p. 82)

2.2. Is there evidence for the presuppositional ambiguity?

Ignoring the syntactic details, let us check more carefully what the analysis can teach us about the issue under consideration here, namely that all indefinite NPs are ambiguous between the presuppositional and existential reading, and that this ambiguity is syntactically encoded. The hypotheses Diesing is suggesting are repeated below, including the one just introduced.

8 a) The restrictive term is presuppositional: Non emptiness of the N set is presupposed iff it occurs in the restrictive term (of a tripartite tree, or formula)

b) The mapping hypothesis
Material from VP is mapped into the nuclear scope
Material from IP is mapped into a restrictive clause.

c) Parameters: English allows LF movement, Dutch and German do not.

The only thing which has been established is the condition on the generic interpretation of bare plurals. The hypothesis which could gain support from this discussion is only (8b). Combining (8b) and (13), we derive that only IP dominated subjects allow the generic interpretation. But nothing follows from that on whether such subjects, or any other IP-subjects are necessarily presuppositional. The issue of presupposition still comes only by stipulation. Diesing's argument can go through without assuming (8a) at all. One can take, instead, a strictly syntactic view of the restrictive term. This, in fact, is the approach in the LF-part of the analysis of Heim (1982, Part 2), where the restrictive term is simply the adjacent sister of the operator. Whether the syntactic restrictive term is also semantically restrictive (or presupposes non emptiness of the N-set) depends, then, on the semantics of the operator. There are several other imaginable ways to encode the insight of (8b) and (13) into the computational system, without ever assuming anything about presuppositions of generic indefinites. Since (as assumed in the analysis following Kratzer) generic indefinites are interpreted pretty much like universally quantified NPs, the question whether their N-set is presupposed is reduced to the general question of presupposition of the universal quantifier. On this question, the two different approaches surveyed in section 1 are in conflict. Diesing's analysis does not shed any new light on this debate (and does not intend to, I think).

Though I see no reason to assume that generic indefinites are ever presuppositional, let us even assume that for the moment. The real leap in Diesing's argument is in concluding from this that all indefinite NPs are ambiguous between a presuppositional and a weak interpretation. E.g. since (19a) is ambiguous between a generic and an existential reading, as we just saw, (19b), which also allows two positions for its subject at LF, is also ambiguous, having a strong (presuppositional) and a weak interpretation.

19 a) Firemen are available.
b) Two/some firemen are available.

There is nothing in the discussion of (19a) from which this conclusion could follow. On the contrary, this discussion takes it for granted that firemen, in the same position, cannot have the "strong existential" reading, which its position presumably allows (see footnote 5 for the details). This does not entail, of course, that (19b) could not have this reading nevertheless. The generic interpretation of indefinites is just irrelevant for the issue under consideration.

Note that the two meaning-distinctions under consideration here are of a very different nature. The generic-existential distinction is a very clear truth-conditional distinction which has always been treated as a real ambiguity. The reason is that it is very easy to imagine a situation in which one of its readings is true, while the other is false. (E.g. if there are 20 fires going on at the moment, and all firemen are busy with these fires, when the 21st call comes, (19a) may be false under its existential reading, since there are no available firemen at the moment, while it is still true under its generic reading.) The ambiguity we are expected to note in (19b), by contrast, is extremely subtle. It can be tested only when the N set is empty, i.e. in a world containing no firemen. Next, what we are asked to check here is our meta-theoretic intuitions regarding whether the sentence is undefined or false, in such a world, or even less reliable intuitions about which contexts we could have uttered the sentence in[50].

Indeed, in contrast to the solid argumentation regarding the generic-existential distinction, the method Diesing uses for showing the presupposition ambiguity is, essentially, declaring that it is there, and providing a textual analysis of imaginary contexts of utterance[51]. Diesing's theory,

---

50

As have often been argued, armchair analysis of imaginary contexts is an extremely dubious procedure, since we may just not be able to think of all possible contexts. Serious discourse researchers such as Ellen Prince or Mira Ariel are usually careful to base their claims on actual discourse examples and statistics, rather than imaginary possibilities.

[51]An example for this procedure is Diesing's discussion of the relevant ambiguity in German. As we saw, German marks the subject position, and Diesing stipulates, further, that the SS position must also be the LF position of the subject. Hence, it follows that sentence (i), which according to Diesing is ambiguous in English, can be disambiguated in German: (ia) with the VP subject is only "existential", while (iib), with an IP subject, is only presuppositional.

i) Two cellists have indeed taken rooms in this hotel.

ii a)...weil ja doch zwei Cellisten in disem Hotel abgestiegen sind.
b)...weil zwei Cellisten ja doch in disem Hotel abgestiegen sind.

To argue that this is the right result, the following highly sensitive and imaginative textual analysis is provided.

thus, makes very fine semantic distinctions, regarding which structures are ambiguous and which are not. But the problem is that there is no obvious way to check them. Specifically, to know whether a sentence (say, (19b) is really ambiguous, there should be some way to disambiguate it. Diesing suggests, in fact, that stress can do the job. Thus, a stressed SOME occurs only with the presuppositional reading. So, in (20) (Diesing's (5b), p. 59), only the presuppositional reading is available, i.e. "it presupposes the existence of ghosts" (ibid).

20) SOME ghosts are in the pantry; the others are in the attic.

21) There are SOME ghosts in the garden.

However, I don't think there is any problem in having the stressed some also in there sentences, as in (21) (and the reader is invited to supply the stories that go along with it). Diesing assumes that the NP position here is only verb-phrasal, which accounts for why only stage level predicates can occur in there sentences (p. 49). Even independently of that, these are the prototypical contexts where we want only weak NPs to be possible. If stressed SOME is a diagnostics of 'strong', or presuppositional, then these contexts allow strong NPs, and we loose the account for why There is every ghost in the garden is out. So, I conclude that whatever the intricacies of SOME may be, its presence is not an indication of the strong reading of the indefinite.

In the absence of a real test for ambiguity, it is not obvious to me what could count as an argument for or against the multiple ambiguities the system generates everywhere. E.g. Diesing argues that an innocent sentence like (22) is three-ways, rather than two ways ambiguous.

22) Every person saw three ghosts.

"In (iia), the cardinal reading is most salient. The sentence asserts the existence of two cellists who have taken room in this hotel." "In (iib), on the other hand... the two cellists are two of some larger set of cellists. The context situation might be one in which a busload of cellists has arrived in town (perhaps for a Villa-Lobos festival) and two of the cellists are staying in this hotel, four more at a local bed-and-breakfast, another with an aunt, and so on. The presuppositional reading is thus associated with the outer, {SPEC IP], subject position." (Diesing, p. 79, her example (42).)

I find highly convincing the claim that (iib) can be uttered in the context described here (which Kuno labelled the 'exhaustive list' use of topic). Questions which come to mind, though, are what are the contexts in which it cannot be uttered, and couldn't (ia) be uttered in a similar context as well. More broadly, it is well known that different discourse contexts correlate with preferred word order, and topic-focus articulation. The question is whether it is the issue of presupposition and previous mention of the set, which affects the preference of word-order here, or some other discourse factors. It may just not be sufficient to imagine one context of possible utterance to answer these questions.

Technically, this is so, since on the narrow scope of the cardinal NP it can still either be attached by QR to IP, or to VP. The following text is designed to show us that this is the right prediction.

"Here the basic context could be visiting an old, rundown house. On the narrow scope, presuppositional reading of three ghosts, the house could be known to be haunted by a set of ghosts such as the ghosts of all Bach's 20-odd children. In this context, every person could see three of the ghosts, the partitive of indicating the presuppositional reading. On the cardinal reading, the house would have no such preestablished reputation. Every person could see three ghosts, even three different ghosts each time. In this case, the sentence in (22) would merely assert the existence of ghosts seen..." (p. 68)

It is hardly possible to disagree that the sentence could be uttered in these different contexts (and many others), or, more generally, that indefinites can be used both when their common noun set has been mentioned already in the discourse, and when not. What is at stake, though, is the question whether this basic fact of language use entails that indefinites are semantically and syntactically ambiguous. If we seriously believed that, then we would have to assume that there sentences are also always ambiguous: I could utter (21) (with or without stress on some) either when I or someone else have already declared belief in the existence of ghosts, or out of the blue.

So far, then, there just seems to be no evidence either for or against the idea of (8a). But, in fact, it is possible to think of some dubious predictions it makes. Diesing assumes that all indefinites, including cardinals, some, etc, are ambiguous in the same way. Next, her analysis of the syntactic difference between stage and individual-level subject is absolute, and independent of the type of the NP determiner. Hence, it predicts that there should be a real semantic difference between (23) and (24)

23) Some/two firemen are available

24) Some/two firemen are brave.

Neither of these allow the generic reading (for independent reasons -this reading is possible only with bare plural or a singular indefinite, see footnote 5). Still, (23) has two LFs, with a VP or an IP subject, while (24) has only the IP-subject structure. It follows from (8a) that (23) should be ambiguous between the existential and the presuppositional interpretation, while (24) is only presuppositional. The question whether (23) is indeed ambiguous is hardly decidable, as we saw. But the question whether there is any difference in number of readings between the two sentences seems a bit more concrete. And it does not seem that (24) allows less interpretations than (23), or that its potential contexts of utterance are more restricted. To the extent that imaginary previous contexts can help here (as the common practice in discussions of presuppositions seems to suggest), I think that (24) can be very easily uttered in

a context like (25), where it is clear that firemen are being mentioned for the first time, so, their existence is not assumed, at least in the established context set.

25) Max: I am doing a research on brave people, but I don't know too many.
Lucie: Well, some firemen are brave, so you may want to check the fire-station next door.

Another way to check the semantic predictions of this analysis comes from symmetry entailments. Under the standard existential analysis, weak NPs are always symmetric. E.g. (a) and (b) below are equivalent to each other. However, the new breed of presuppositional indefinites can no longer have this property. This may be illustrated in (26). If some ghosts is presuppositional, as in Diesing's "strong interpretation", then (26a) is undefined (assuming ours is a ghostless world), while (26b) is false. Of course, there is no way to check our intuitions on this in the case of (26), since Diesing's point is precisely that (26a) is ambiguous, so it is consistent with either decision, depending on the structure we choose for it. However, the system entails that this is not so in (27). Since the predicate is individual-level, the subject must be in IP position. Hence it has only the presuppositional interpretation. Therefore, (27a) and (27b) cannot be equivalent.

26 a) Some ghosts study Dutch
b) Some students of Dutch are ghosts.

27 a) Some ghosts are speakers of Dutch
b) Some speakers of Dutch are ghosts

Again, there seem to be no difference in the number of readings here. For someone like myself who sees the sentences in (26) as equivalent (and, in fact, has difficulties imagining what others could mean when they don't), the sentences of (27) are equally equivalent.

3. Topics and discourse anaphora.

1. The insufficiency of familiarity

One of the areas where topics play a crucial role is discourse anaphora. This is also an area in which an impressive amount of wrong claims has been made over the years.

Kuno (1972), (1975) argued that backwards anaphora is possible in discourse only when the antecedent has been previously mentioned (1972), or is "determinable (predictable) from the preceding context" (1975). This, apparently, is a very appealing claim. It feels 'intuitively right', which may explain why it keeps being rediscovered and proposed periodically by syntacticians working on anaphora.

But how could we test such an hypothesis? Kuno used introspection - a common practice in arm-chair discourse theory. Note, however, that this is not a claim about syntax, but about the use of sentences in discourse. There is, therefore, a very simple way to check it, namely look at actual occurrences of backward anaphora in discourse. Carden (1982) has done precisely that: He collected a corpus of examples from written texts, and the results are unequivocal: In well over 50% of the examples the antecedent is mentioned for the first time. In many of the cases the antecedent is indefinite, so there is no issue of previous mention. Here are two of his examples.

1) "Because they wanted to know more about the ocean's current, students in the science club at Mark Twain Junior High School of Coney Island gave ten bottles with return address card inside to crewmen of the NY city's sludge barges" (The New York Times)

2) "When she was five years old, a child of my acquaintance announced a theory that she was inhabited by rabbits". (New York Times)

(The same results were found recently in a statistical study of Dutch written discourse by Mulders and Jas (1995), where around 60% of the occurrences involved antecedents which are brand new discourse entities.) Carden concludes his findings with a warning which is well worth remembering: It is not possible to establish discourse generalizations based on mere introspection. The least we should expect is that actual discourse will provide some confirmation to our hypotheses[52].

---

[52]The problem with the introspection method in discourse is that we never know precisely what factor is the one relevant for our judgments of our invented examples. E.g. Kuno (1972) supports his generalization with the following invented discourse: (Underlining indicates coreference).

i a) What do you suppose I did when I saw Harry getting mad?
b) ?-I calmed him before Harry did something rash.

Methodology aside, this little fragment of linguistic history, may shed some light on an on going debate regarding the role of familiarity and previous-discourse mention in discourse theory. I mentioned already one area of this debate. In the research of topic, two descriptions have been proposed since very early on: The one is that NP-topics the discourse entities that the sentence asserts something about. The other is that NP-topics are the old information or given entities of the discourse. The first is vague, and difficult to explain. The second is a purely quantitative concept which seems clear and easy to check. It is probably for this reason that most discourse studies adopted the second. But, as I mentioned, in Reinhart (1981), I argued in detail that this simpler and cleaner line is, unfortunately, empirically wrong. Topics cannot be defined as given, previously mentioned, discourse entities, and there is no way around attempting a serious definition of aboutness. Following the insight of Stalnaker's (1978) views of the context set, rather than defining topics in terms of the past -the effect of previous discourse on the given sentence, the analysis I attempted there is coached in terms of the future -their effect on the ongoing discourse.

More generally, I argued that familiarity, and other notions of the discourse past, play less direct a role in discourse analysis than commonly assumed. Fifteen years later, in view of the findings in discourse theory, I believe this claim even more. The question whether an entity (or the set it belongs to) has been mentioned already, does play a role, e.g. as one of the factors in identifying topics, but it has no independent theoretical role in and of itself. Specifically, there is no coded feature of language that is determined just by the question of previous mention.

Let us take a very simple question a linguist may want to ask: What are the conditions under which a definite description is used in discourse, i.e. when do speakers refer to discourse entities with definite descriptions. Heim (1982) offers the following familiarity-based generalization:

3) A definite is used to refer to something that is already familiar at the current stage of the conversation. (Heim, 1982, (1), p. 164)

(definite is used in (3) to cover both definite descriptions and pronouns. But I focus here only on the first.) This is only the informal statement of the generalization. In its formal statement

---

ii a) Who do you suppose I calmed when I saw him getting mad?
b) *-I calmed him before Harry did something rash.

Presumably, (iib) is bad, but (iia) is better, because the antecedent Harry is previously mentioned. The truth is that the occurrence of (b) in both these contexts is pretty weird. But all this means is that we don't know exactly what are the conditions that make backward anaphora perfect in the real discourse examples. It is easy, therefore, when we look at bad invented discourses to convince ourselves that the one which confirms our theory is slightly less bad.

((15), p. 175), Heim clarifies the notion of familiarity, and allows a definite description to be used only if its index is already in the context set (file). i.e. if it was previously mentioned[53].

But (as has been widely observed), this is not the correct generalization governing the use of definite descriptions, and they are frequently found also with no previous discourse mention (The president, the man with the champagne glass, the winner of the race, my neighbor, etc.) A way to approach this problem is to attempt a different definition of the concept of familiarity. Indeed, at the same period (of early 80's), more sophisticated descriptions were discovered in discourse theory, particularly in the work of Prince (1979, 1981), which is based on a careful textual analysis of actual discourse.

Abstracting away from many important details, the generalization which emerged from these discourse studies at the time can be summarized as follows. Corresponding to the three types of familiarity of discourse entities, below, there are different anaphoric expressions, that mark the different entities:

4)

familiarity hierarchy	**Linguistic markers**
textually given entities	pronouns
situationally given entities	demonstratives
world knowledge	definite descriptions

The familiarity hierarchy ranks the entities from the most familiar, or readily available, (textually given, i.e. entities mentioned already in the previous discourse) to the less available, but still familiar ones (world knowledge). Following an important modification by Ariel (1990), we may assume, further, that an expression marking the less available entities can also mark more available ones, but not conversely. Then the result is that a definite description can mark entities which are either given in our shared world knowledge (the president), in the situation of utterance (the man with the champagne glass), or the previous discourse. But (destressed) pronouns can only refer to the later entities.

---

[53] The relevant part of Heim's formulation is given in (i). (I am only discussing here the aspects of her analysis that regard the use of definites.)

i) Let F be a file, p an atomic proposition. Then p is appropriate with respect to F only if, for every noun phrase NP_i with index i that p contains:
if NP_i is definite, then i c DOM (F).
(Heim, 1982, (15), p. 175)

[HAND: c = MEMBER]

This takes us much closer to a description of the conditions of use of definite descriptions (though, as argued by Ariel (1990), it is nevertheless not the correct description). But it does not yet answer the question we posed. The question was when such descriptions would be actually used. Specifically, since it can refer to all kinds of entities on the familiarity scale, how do we know when to use a definite description and when to use a pronoun. This, in fact, was just an initial illustration of the problem I want to survey here. The real issue is how do we actually resolve anaphora in discourse. For this, familiarity theory is not sufficient.

2. Accessibility theory[54]

In a pioneering work, Ariel (1990) points out that while the correlation observed in the table above may be interesting and partially true, it cannot take us very far on the issue of actual anaphora resolution in discourse. If we focus on textually given entities, we find that, in fact, any of the linguistic markers above can refer to such entities. In any actual discourse (written text or oral discourse), it is always the case that there are many textually given entities, (in DRT terms there are many files, or card-entries we have already opened) and the question is how we refer back to them, or which expressions refer to which entity. Usually we refer back to discourse entities with reduced, or anaphoric expressions that do not carry the full information. So the problem is deciding which is the relevant antecedent. It almost seems surprising that speakers, in fact, do not have any problem with this task.

One invented example that illustrates this is (5):

5) Max was walking down from school, pondering about the meaning of life. Soon he ran into Felix and he suggested that they stop at the bar. (Did Max or Felix suggest the bar?)

6) ...Al snel kwam hij Felix tegen en
   a.hij
   b.deze suggereerde dat ze naar een bar zouden gaan

The pronoun he can refer to either of the given discourse entities Max or Felix. In fact, most speakers (e.g. when we checked this in class), identified he as referring to Max. Had we wanted the pronoun to refer to Felix, it would have to be stressed. The situation is even more interesting in the Dutch example in (6), where the pronoun (6a) refers for most speakers to Max, but the demonstrative deze in (6b) refers unequivocally to Felix.

[54]Accessibility theory was proposed in Mira Ariel's Tel Aviv dissertation, 1987, and further developed in her (1990) book.

The intuitions here are very sharp, (unlike, say some of the intuitions on what 'specific' is, or other discourse notions we examined). Next it is clear that there is some relation here between the linguistic form of the expression and the assignment of value. So this is a clear case of a relation of form and context, which should be captured by the theory of interface.

Ariel's search for the factors effecting the choice of antecedent started with a textual count of the distribution of the various types of expressions. Her first finding is that the distance between previous mention of the entity and the anaphoric expression plays a crucial role. On this count she found that pronouns are most common where the entity they refer to was mentioned in the same sentence or in the preceding one. Definite descriptions are most common when the antecedent is further in the same paragraph or in the previous paragraph, and demonstratives chose antecedents in between. (Table 1 p. 18). Lets assume that the nearer the antecedent is the more accessible it is. What we get, then, is that pronouns signal that the antecedent is very accessible, demonstrative instruct us to choose a less accessible antecedent, and the definite description signals an even less accessible antecedent.

It should be obvious, though, that this cannot be sufficient. E.g. it does not take us very far in explaining (5). Both the entities Felix and Max are mentioned in the sentence immediately preceding the pronoun, so they must be equally accessible (or, if literal distance counts, Felix is nearer). Still the pronoun prefers Max, and the demonstrative - Felix.[55] Ariel argues, further, that there is another factor determining accessibility: topichood. The topic is always highly accessible. (So far we are using this term in an intuitive way, and we will return to a more precise definition latter. Ariel provides ample substantiation for this claim. In fact, much of the discourse work on anaphora resolution has found the same results - see the reference cited by Ariel.) In (5) Max is the discourse topic. (We have been talking about him and his mental state, etc.), hence even though the distance criterion does not select between this candidate for the value of <u>he</u> and the other, the topichood criterion signals Max as the most accessible discourse entity.

The picture is, then, summarized in (7) and (8). The degree of accessibility of an entity (antecedent) is determined by distance and topichood. (Further elaboration is needed on the relative weight of the two criteria. Generally, as long as the same topic is kept, we can refer back to it with a pronoun even far away. It is only if there are intervening topics, that the distance plays a role). A third accessibility factor, which Ariel does not discuss, is reference to the center. This is needed to allow accessibility theory to account correctly for the distribution of logophors, but I cannot discuss it here. The different linguistic forms of anaphoric

[55]ELABORATE? Indeed Ariel notes that this correlate with another puzzling fact. If we look only at the distribution of pronouns, we still find many pronouns referring far back in the discourse. The next factor is topichood. If we take out all pronouns referring to the discourse topic, we find a much more local distribution of pronouns. (tables 2, 3, p. 19)

expressions are markers of the degree of accessibility of the antecedent - they signal where, in the discourse storage, should the antecedent be retrieved from. So far we assumed only three such markers, but, in fact, these are only representatives of the type of expression in each accessibility rank. A more detailed list is given in (8), and Ariel (and others) argue that there is also further internal ranking. (We may view the list as just a continuum marking varying degrees of accessibility.)

Accessibility of Discourse-antecedent:

7) a. Distance: The closer the antecedent is, the more accessible it is.
   b. Topichood: Topic antecedents are always highly accessible.

======

   c. Center: Discourse - participants (speaker, center of consciousness) are highly accessible (Not discussed by Ariel)

Accessibility markers (Partial list):

8) i. High accessibility: anaphors > clitic and 0 pronouns > pronouns >
   ii. Intermediate accessibility: stressed pronouns> demonstratives (this >that) >
   iii. Law accessibility: Definite descriptions (including epithets) > names.

The psychological reality of the hierarchy in (8) has been witnessed in many studies of anaphora resolution (see references in Ariel). E.g. in reading time experiments by Purkiss (1978), it took subjects longer to read and process the text in (9) with it, than with the baby. This is so, since the discourse topic is the mother. The baby is not as highly accessible discourse entity as the choice of the high accessibility marker pronoun for it suggests. This choice, then, does not result in optimal discourse. (reported in Ariel, p. 23)

9) Context:
   The mother picked up the baby. She had been ironing all afternoon.
   She was very tired.
   Target:
   It/ the baby had been crying all day.

10) a. First square 19 and then cube it (it = 19 or 19^2)
    b. First square 19 and then cube that (that = only 19^2)
    (Isard (1975), reported in Ariel, p. 25)

    c) First multiply 4 by 3 and then divide it/that by 2.

Another illustration is Isard's findings in (10a,b). The number actually mentioned is a more highly accessible discourse antecedent than the number which has to be calculated (both since the first is a topic, and since explicitly mentioned entities are more accessible than inferred ones). Avoiding a pronoun, then, as in (10b) signals to the hearer that the less accessible antecedent must be selected. Similarly in (10c), when that was used, the result was

unanimously 6 (rather than 2).[56]

The intuition behind accessibility theory, then, is that the various forms of anaphoric expressions that natural language allows play a role at the interface. It is useful to have so many anaphoric forms, since they can be used to identify different antecedents. The syntax determines only when an anaphoric expression is permitted. When syntax allows more than one (which is usually the case), the choice between the different permitted forms is determined by discourse needs at the interface.

We may note that the type of reasoning assumed in the computation of accessibility is reminiscent of the optimality strategy which has gained much theoretic attention recently, mainly in phonology, but extended to other areas in work by Grimshaw and Pesetsky. The idea is that (unless explicitly prohibited by the computational system), whether a given anaphoric expression is appropriate in a derivation is determined by the type of competitors, and the optimal choice is always relative to the alternative options. This line of reasoning would be clearer with an example:

[56] Several anaphora puzzles are explained by accessibility. E.g. The intuition that in (ib) only Lili is the antecedent of the pronoun is quite strong, though it cannot follow from any sentence level constraint.

i) a. When she entered the room, Lucie greeted Lili. (she = Lucie or Lili)
   b. When she entered the room, Lili was GREETED by Lucie. (She = Lili)

In Reinhart (1981) it was argued that backward anaphora is possible only when the antecedent is the topic. Passive subjects, unlike regular subjects, are always topics. Hence, while in (ia) there is no way to know what the topic is, in (ib) it is Lili. (The focus is on the verb, so that whatever is involved here is an issue of topics, not foci)

Dutch zijn and diens:

The pronominal system of Dutch has a clear distinction between high and intermediate accessibility markers: While pronouns, like hij mark high accessibility, the demonstrative deze marks intermediate accessibility. This was illustrated in example (6). We should note that deze is used quite frequently in written discourse, with no air of markedness. Class participants have found tens of examples in both newspapers and literary texts. All confirm to the generalization that deze is used for less accessible antecedents, and never refers back to the current topic.

As further illustration of the accessibility analysis, I would like to examine in detail a slightly more complex case: Dutch has two possessive pronouns - zijn and (the demonstrative) diens. Unlike the standard demonstrative (deze), the possessive diens is used more rarely. It is stilted, restricted to written discourse, and is used to refer only to 'famous' antecedents. Still, when allowed, it appears to obey surprising restrictions. At first glance, these appear to be syntactic restrictions, as, indeed, was argued by Postma (1984).

Postma observed that in (9) diens can take only the non-subject lubbers as its antecedent.

9) a. Lubbers begroet Mitterand bij zijn aankomst op het vliegveld (zijn = Lubbers or Mitterand)
   b. Lubbers begroet Mitterand bij diens aankomst op het vliegveld (diens = Mitterand)
   Lubbers welcomed Mitterand on his (zijn/diens) arrival at the airport. (Postma's example 1).

10) De vader van Jan heeft diens boeken weggegooid
   The father of Jan has thrown away his books (diens=Jan)

Postma concludes that the distribution of diens obeys a syntactic restriction, which is roughly, that it cannot be coindexed with a subject. This is further illustrated in (10). Furthermore, diens is a pronoun that allows only coreference, (rather than variable binding), but as such it marks obligatory coreference, hence cannot be used without an antecedent. These two assumptions combined, entail that the sentences in (11) are out, since there is no potential antecedent[57].

[57] An exception, noted by Postma is ECM subjects, as in (i).

(i) Jan zag [Piet in diens tuin lopen] (diens = Piet)
   Jan saw [Piet walk in his garden]

11) a. *Assad overwon diens nededinger
Assad conquered his rival
b. *Jan zei dat Piet diens boeken moest verkopen
Jan said that Piet had to sell his books.

Syntactically, this is a puzzling conclusion, since it requires a special stipulation prohibiting coindexation with subjects. Let us examine, next, the behavior of diens with respect to accessibility. In a context as (5), repeated with a slight variation, we see that the choice of diens functions the way deze did: In (6b), the son is unequivocally Felix' rather than Max' son. Given that Max has been the discourse topic, this suggests that diens is a demonstrative, wrt accessibility, signalling that the antecedent should be the less accessible one.

5) Max was walking down from school, pondering about the meaning of life. Soon he ran into Felix and his son suggested that they stop at the bar.

6) ...Al snel kwam hij Felix tegen en
a.zijn zoon
b.diens zoon suggereerde dat ze naar een bar zouden gaan

The question, then, is whether an accessibility account can replace the problematic syntactic account. In actual discourse it is often the case that entities mentioned in subject position are the most accessible entities (since this is often a topic position, and, as observed by Prince, there is a proven tendency in discourse to place accessible entities in subject position). To see whether it is the accessibility factor or subjecthood per se which is relevant here, we should find an example where the entity in subject position is nevertheless not the most highly accessible one. (12) (from Ad Neeleman, p.c.) is a case in point.

12) Lubbers is de laatste tijd zichzelf niet meer. Gelukkig merke Mitterand niet op dat hij in slaap viel tijdens diens toespraak.
(Lubbers is not himself lately. Fortunately, Mitterand didn't notice, that he fell asleep during his-diens speech)

diens is acceptable here, referring to Mitterand. Though the antecedent -Mitterand is in subject position, the discourse makes it clear that Lubbers is the discourse topic.

The syntactic generalization, thus, can be best understood as a mere indirect reflection of the accessibility strategy. Given no actual discourse, as in the case of (9) and (10), the hearer attempts to justify the use of diens, looking for the least accessible antecedent. Given no other way to decide, he would go for the non-subject, which is generally the less accessible candidate. When there is no more than one candidate, as in (11a), the sentence looks weird, since there is no possible reason to use the marked expression diens when there is no competition. Postma's judgment of (11b) remains unexplained, and in fact, speakers in class thought that it is not so bad with the embedded subject as the preferred antecedent[58].

[58]Another part of Postma's analysis is the observation that diens does not allow bound-variable interpretation. This requires further study. It has often been claimed that the lower accessibility markers do not allow this interpretation. (E.g. it was claimed about kare in Japanese). If true, this requires an independent explanation. E.g. Tohru Noguchi, in a recent Amherst dissertation argues that only determiners in D-position can be bound variables, while demonstratives are usually N-pronouns. (A similar distinction is also assumed by Joost Zwartz (1993).

3. Epilogue: Anaphora as test for topichood

One of the factors that make topics a harder subject for research than, say, foci, is that in most languages they are not marked either intonationally or syntactically. They have to be computed directly at the conceptual interface. For this reason, the issue of how do we know whether our theory identifies correctly the topics is a serious one. Several attempts keep popping up, periodically, to identify topics with a syntactic position (in non topic languages), particularly with the initial position of the sentence. (This idea, originally due to the Prague school, is recently advocated again by Vallduvi (1992).) This has been debated alot, and rejected in the past (see Reinhart (1981), Gundell (1974)). Other things being equal, a c-commanding NP is preferred as topic over NPs it c-commands, but there are other factors that can undermine that. So, the issue of tests is a serious one. Although I offered some such tests in Reinhart (1981), and others did as well, I think these are not sufficiently decisive.

However, accessiblility thoery may just be the clue we needed. Since anaphora resolution is a concrete fact, which speakers perform with amazing unanimosity, and since it is sensitive to factors like topichood, we can use anaphora resolution as a test. Specifically, if our hypotheses define a certain NP as a topic of a given sentence in a given context, the prediction is that it should not be possible to refer back to this NP with a low accessibility marker like a demonstrative. This proved particularly useful in languages like Dutch or German that have the active demonstrative deze for such purposes. In English, this test has only a limited scope, since we can only use stressed pronouns, which involve additional complexities independent of accessibility.

Let me illustrate one area where this test has prooved helpful in checking an hypothesis. Unlike the case of subject scrambling discussed in section 1 of this part, the distribution of object scrambling in Dutch cannot be explained in terms if topics, since the scrambled NP need not be a topic. This is witnessed by the fact that it can be referred back to by deze, as in (13), from Ad Neeleman. (This can be compared to examples (6) and (9) above, where deze cannot refer to the topic.)

13) Lubbers zal [een niet met name genoemde Europese premier]$_i$ morgen op het vliegveld ontmoeten, wanneer deze$_i$ uit Tel Aviv terugkomt

Lubbers will a not by name mentioned prime minister tomorrow at the airport meet, when this one from Tel Aviv returns

This is among the reasons for why I argued (in the focus part III) that object scrambling is explained by focus, rather than topic considerations.

References for Part III

Chomsky, N. (1971), "Deep structure, surface structure and semantic interpretation" in D. Steinberg and L. Jakobovits (eds) Semantics, An Interdisciplinary Reader In PHilosophy, Lingustics and Psychology, Cambridge University Press.
Chomsky, N. (1976) "Conditions on rules of Grammar" Linguistic Analysis, vol 2, no. 4. Reprinted in N. Chomsky, Essays on Form and Interpretation, North Holland, Amsterdam, 1977.

Chomsky, N. and M. Halle (1968), The Sound Pattern of English, New York: Harper and Row.
Cinque, Guglielmo (1993) "A null theory of phrase and compound stress", Linguistic Inquiry vol 24, no 2; 239-298.
Delfitto and D'hulst (1994), "Beyond the mapping hypothesis" ms. U of Utrecht,
Enc, M. (1991)) "The semantics of specificity" Linguistic Inquiry 22, P. 1-25.
Gussenhoven, C. (1984) On the grammar and semantics of sentence accents in Dutch, Foris
Halle, M. and J.R. Vergnaud (1987), An Essay on Stress. Cambridge, Mass: MIT Press.
de Hoop, H. (1992) Case configuration and NP interpretation, PhD dissertation, Groningen)
Jackendoff, R. (1972), Semantic Interpretation in Generative Grammar, Cambridge, Mass: MIT Press
Ladd, D.R. (1980), The Structure of Intonational Meaning Bloomington: Indiana University Press.
Pinto, M. (1994) "Subjects in Italian: Distribution and interpretation", ms. University of Utrecht, OTS
Neeleman, A. (1994) Complex Predicates, PhD dissertation, Utrecht University/OTS
Reinhart, T. (1981), "Pragmatics and linguistics: An analysis of sentence topics", Philosophica volume 27, no 1. Distributed also by Indianal University Linguistics Club, Bloomington, Indiana.
Selkirk, E. (1984), Phonology and Syntax: The relation between sound and structure. Cambridge, Mass: MIT Press
Truckenbrodt (1993)
Rooth, M. (1992) "A theory of focus interpretation" Natural Language Semantics, vol 1: 75-116
Vallduvi, E. (1990) The Informational Component PhD dissertation, University of Pennsylvania
Vallduvi, Engdahl, and Herman Hendricks.)
Zubizarreta, Maria Luisa, (1994) "Word order, prosody, and focus", ms. U of Southern California.

References for Part IV

Ariel, M. (1990), Accessing Noun Phrase Antecedents, London and New York: Routledge.
Barwise, J. and R. Cooper (1981), "Generalized quantifiers and natural language", Linguistics and Philosophy 4, 159-219.
Carden, G (1982), "Backward anaphora in discourse context", Journal of Linguistics, 18, p. 361-387
Diesing, M. (1992) Indefinites, MIT Press, Cambridge, Mass
Fodor, J. (1979), "In defence of the truth gap", in C.K. OH and D. A. Dinneen (eds) Presuppositions, Syntax and Semantics 11, New York: Academic Press.
Heim, I. (1982a), "File change semantics and the familiarity theory of definiteness" in R. Bauerle eta al, eds. Meaning, use and the interpretation of language. Berlin/New York: de Gruyter, p. 164-189.
Heim, I. (1982b), The semantics of definite and indefinite Noun Phrases,PhD dissertation, University

of Massachusetts, Amherst, Published in 1989 by Garland, New York.
Gamut (1991) Logic, language and meaning, volume 1, University of Chicago Press.
Goodman, N. (1972), "About" in Goodman, N., Problems and Projects, Indianapolis
Groenendijk, J and M. Stokhof (1984) Studies on the semantics of questions and the pragmatics of answers PhD dissertation, University of Amsterdam.
Gundel, J. (1974), The role of topic and comment in linguistic theory, PhD dissertation, University of Texas at Austin, distributed by Indiana University Linguistic Club, Bloomington, Indiana.
De Jong, F. and H. Verkuyl (1985), "Generalized quantifiers: The properness of their strength", in J. van Bentham and A. ter Meulen (eds) Generalized Quantifiers: Theory and Application, GRASS 4, Dordrecht: Foris.
Keenan, E. (1987), "A semantic definition of indefinite Nps", in ter Meulen and E. Reuland (1987)
Kratzer, A. (1989) "Stage and individual level predicates" in Papers on Quantification. NSF Grant report, Department of Linguistics, U Mass Amherst.
Kuno, S. (1972), "Functional sentence perspective: a case study from Japanese and English" Linguistic Inquiry, 3, p. 269-336
Kuno, S. (1975), "Three perspectives in the functional approach to syntax", Papers form the parasession on functionalism, Chicago linguistic Society, p. 276-336.
Lappin, S. and T. Reinhart (1988) "Presuppositional effects of strong determiners" Linguistics vol 26:1021-1037
ter Meulen, A. and E. Reuland, eds (1987) The Representation of (In)definites, Cambridge, Mass: MIT Press.
Mulders, Iris and Floor Jas (1995) "They do exist: Backward anaphors in Dutch" Manuscript, University of Utrecht.
Prince, E. (1979), "On the given/new distinction," in W. Hanks, C. Hofbauer and P. Clyne (eds) Papers from the fifteenth regional meeting of the Chicago Linguistic Society, p.267-78
Prince, E. (1981), "Towards a taxonomy of Given-New information", in P. Cole (ed.) Radical Pragmatics, New York: Academic Press, p. 233-255
Putnam, H. (1958) "Formalization of the concept 'about'", Philosophy of science vol 25, no 2.
Postma, G (1984), "The Dutch pronoun diens; distribution and reference properties", in Hans Bennis and van Lessen Kloeke (eds) Linguistics in the Netherlands.
Reinhart, T. (1981), "Pragmatics and linguistics: An analysis of sentence topics", Philosophica volume 27, no 1. Distributed also by Indiana University Linguistics Club, Bloomington, Indiana.
Reinhart, T. (1987), "Specifiers and operator binding" in ter Meulen and Reuland (1987).
Strawson, P. (1964), "Identifying reference and truth values", Theoria vol 30. Reprinted in Strawson, P. Logico Linguistic Papers, London Methuen (1974), and in D. Steinberg and L. Jakobovits (eds), Semantics London: Cambridge University Press.