Derivations & Evaluations. On the syntax of subjects and complementizers
Dekkers, J.R.M.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
2. Derivations & Evaluations

0. Introduction

In the early 1990s, Noam Chomsky launched his Minimalist Program (published in Chomsky, 1995). The main objective of this program is to reconsider which notions are really indispensable in a theory of language. Of course, minimal redundancy as such is not worthy of a special program; it should be a structural target in linguistic inquiry. However, the justification of the program is found not so much in the goal of a perfect model of language itself, but rather in the rigorous way Chomsky tries to reach it.

The Minimalist Program should be considered a reaction against the Principles & Parameters (P&P) framework which dominated the eighties. With four levels of representation, phrase markers of immense proportions, complicated conditions on both the application of transformations and the occurrence of their traces, ad-hoc definitions of auxiliary notions such as government and command, a proliferation of indices, and a host of empty categories, P&P was an empirically powerful, but formally intricate model of syntax. In the Minimalist Program, Chomsky frees us in one step from most of the aforementioned notions. The starting point is minimal redundancy, instead of a maximal empirical adequacy:

[We seek to determine just how far the evidence really carries us toward attributing specific structure to the language faculty, requiring that every departure from "perfection" be closely analyzed and well motivated. (Chomsky, 1995: 9)]

The term Minimalist Program is often used to refer to the result of Chomsky’s recent work. However, we will use it in its original, more dynamic meaning, and propose several important modifications of the model of syntax put forth in Chomsky (1995) that lead to a simpler model or an increase in empirical adequacy.

The main characteristics of Chomsky’s (1995) model of syntax are sketched in section 1. Section 2 focuses on economy of derivation. We will propose a unification of two of the main economy conditions as well as a reformulation of feature strength in terms of optimality-theoretic constraint interaction. The empirical advantages of this step will be discussed in section 3. This leads us to a model of syntax whereby representations are derived in an impoverished Chomsky-style computational system, and evaluated as in OT. Further properties of this Derivations & Evaluations framework, which is originally proposed in Broekhuis & Dekkers (to appear), are presented in section 4.

1. The computational system

This section presents the syntactic system proposed by Chomsky (1995). We will focus on the reduction of levels of representation (section 1.1), on the motivation for movement (section 1.2), and on structural locality (section 1.3).
1.1. Levels of representation

Since the late 1970s, it has been assumed that grammar consists of four levels of representation: D-Structure, S-Structure, Logical Form, and Phonetic Form. These levels are organized in a T-shaped system:

![Diagram 1: T-shaped system of representation levels]

According to Chomsky (1995), however, a two-level system should suffice, since the task of grammar is to pair up sound and meaning. In his terminology, one level is needed to feed the articulatory-perceptual (A-P) system, while the other provides the conceptual-intentional (C-I) system with instructions. This gives the grammar in figure 2.

![Diagram 2: Two-level system of representation]

The computational system producing LF representations (henceforth the computational system or $C_{HL}$) consists of three operations. The operation Select takes elements from the numeration. The numeration ($N$) is a set of lexical items arbitrarily put together. Each member of $N$ must be selected exactly once. Merge creates structures by assembling the selected members of $N$. Move transforms these structures. $C_{HL}$ targets a syntactic representation that is interpretable by the C-I system. Since grammar relates sound and meaning, the LF module must feed the PF module at some point (Spell-Out). The representation present at the point of Spell-Out should not in any way be considered a level of representation comparable to S-Structure. Spell-Out feeds the PF component. This component transforms spelled out structures so that they can be interpreted by the A-P system.

Since the minimal set-up of grammar in figure 2 lacks D-Structure and S-Structure, the two internal levels of representation in figure 1, we should wonder whether the conditions and principles that hold at these levels in the P&P model can be reformulated so that they apply either at LF or at an arbitrary point in the derivation, in accordance with figure 2.

Let us start with D-Structure. In P&P, D-Structure functions as the interface between the lexicon and the computational system. It represents arrays of lexical
items in accordance with X-bar Theory (or with a less uniform set of re-write rules in earlier stages of the theory). Crucially, D-Structure precedes the transformational component that gives us S-Structure and Logical Form. Chomsky (1995) advocates an alternative approach, whereby phrase structure is built up transformationally, i.e. by the application of the aforementioned Merge. This operation successively transforms pairs of syntactic objects into larger structures. The operation Move is defined along the same lines as Merge, with the additional condition that one of the merged objects originate from inside the other object. Thus, Chomsky returns to the generalized transformations he postulates in his early work (see Chomsky, 1975):

(1) a. Merge: Take a pair of syntactic objects \((\alpha, \beta)\) and replace them by a new combined syntactic object \(\gamma\).
   b. Move: Take a pair of syntactic objects \((\alpha, \beta)\), \(\alpha\) contained in \(\beta\), and replace them by a new combined syntactic object \(\gamma\).

Contrary to the re-write rules responsible for generating D-Structure, Merge does not necessarily apply before Move. This has several positive side-effects. Let us, for instance, take the sentences in (2). In (2a), John and he cannot be co-referential, while in (2b) they can. Following Lebeaux (1988), Chomsky argues that this asymmetry is due the fact that the relative clause in (2b), being an adjunct, may be inserted after movement of the \(wh\)-phrase, while the complement clause in (2a) must be inserted before \(wh\)-movement has applied. He further supposes that movement transformations leave behind traces in the form of a copy of the moved element. This gives the LF-representations in (3). Now, (3a) violates Condition C of the Binding Theory (see Chomsky, 1981; see also chapter 4) because John is c-commanded by the coindexed he. In (3b), on the other hand, no such violation takes place because the clause containing John has been inserted after the \(wh\)-phrase has moved from the position where it is c-commanded by he.

(2) a. Which claim that John was asleep was he willing to discuss?
   b. Which claim that John made was he willing to discuss?

(3) a. [which claim [that John was asleep]] was he willing to discuss [which claim [that John was asleep]]
   b. [which claim [that John made]] was he willing to discuss [which claim]

Let us now turn to S-Structure. Chomsky makes a distinction between two classes of potential evidence in favor of this internal level of representation:

(4) a. There are syntactic principles that apply at S-Structure.
   b. In some languages the operation \(\alpha\) applies overtly, while in others \(\alpha\) applies covertly.

Let us start with (4a). Clearly, we cannot dispense with S-Structure if there are syntactic principles that refer to it. The three most likely candidates are Case Theory, Binding Theory, and Bounding Theory. Since our minimal syntax is of a highly derivational nature and disposes of one syntactic interface (LF), these conditions should be re-interpreted as conditions that apply at an arbitrary point in
the derivation or at LF. It seems that this is indeed possible. In this section, we limit ourselves to the Binding Theory. Case Theory is discussed in section 1.2 below, in conjunction with (4b). We will briefly address Bounding Theory in section 4.3.

There are reasons to take the Binding Theory not to hold at S-Structure, but rather at LF. Broekhuis & den Dikken (1993) give the examples in (5) (which are not mentioned in Chomsky, 1995). Example (5a) is compatible with the idea that (Principle A of) the Binding Theory applies at S-Structure. Bill c-commands himself and is in the binding domain of this anaphor. Therefore, Bill binds himself. John, on the other hand, does not bind himself because it is not contained in the binding domain.¹ In (5b), himself is c-commanded by John, but not by Bill. Hence, it is incorrectly predicted that himself can only be bound by John. This suggests that Principle A can also be satisfied before the NP containing himself moves to its scope position.

(5)  
   a. John does not know that Bill saw a picture of himself.⁵
   b. John does not know which picture of himself Bill saw.

Remember that we are assuming that movement leaves a copy. This means that the LF representations of (5) are as given in (6). In (6b), which picture of himself undergoes movement to the lower SpecCP, and leaves a copy in its Case position. In (6a), a picture of himself does not undergo A'-movement.

(6)  
   a. John does not know that Bill saw [a picture of himself]
   b. John does not know [which picture of himself] Bill saw [which picture of himself]

Chomsky (1995) argues that Principle A is satisfied if some copy of the anaphor is locally bound at LF.² In (6b), John can bind the anaphor contained in the lower SpecCP and Bill can bind the anaphor dominated by the complement of the lower verb. A binding ambiguity is the result. In (6a), on the other hand, only one copy is available, which can only be bound by Bill.³ Hence, no ambiguity arises.

One could argue instead that we have the choice of applying binding conditions either at D-Structure or at S-Structure. In (5a), himself is bound by Bill at both D-Structure and S-Structure. The structures are given in (7).

---

¹ If himself is contained in an NP that occupies the object position of the lower clause, as in (5a), its binding domain (governing category) is the lower clause (or rather, the lower IP). If, on the other hand, the NP occupies the lower SpecCP, as in (5b), the binding domain is the higher clause (or rather, the higher IP). See Chomsky (1981) for details.
² For expository reasons, we will ignore some of the technical details of Chomsky’s analysis, which include Quantifier Raising and deletion.
³ Vanden Wijngaard (1994) argues that himself is not necessarily interpreted as an anaphor. This allows him to postulate that Binding Theory invariably applies to elements in their base position. However, if this were correct, we would incorrectly predict that (5a) is ambiguous on a par with (5b).
a. **D-Structure:**
John does not know that Bill, saw [a picture of himself]
b. **S-Structure:**
John does not know that Bill, saw [a picture of himself]

For (5b), on the other hand, D-Structure and S-Structure are distinct, as shown in (8). A binding ambiguity is the result: if we choose to apply Principle A at D-Structure, Bill binds himself, while at S-Structure, John binds himself.

(8) a. **D-Structure:**
John does not know Bill, saw [which picture of himself]
b. **S-Structure:**
John; does not know [which picture of himself;] Bill saw /

However, this leads to the prediction that anaphors can only be bound in the Case or the scope position of the NP they are contained in. As noted by Broekhuis & Den Dikken (1993), this prediction is falsified by the binding ambiguity in (9). This example suggests that himself can also be bound in an intermediate position.

(9) Which picture of himself doesn’t John know that Bill saw?

If movement applies successive cyclically (see section 4.3 below), which picture of himself moves from its Case position to the higher SpecCP via the lower SpecCP, as in (10). This means that both the Case position and the intermediate SpecCP contain a copy. Like in (6b), John binds the intermediate occurrence of himself whereas Bill binds the lower occurrence. Hence, (9) is correctly predicted to be ambiguous by Chomsky (1995).

(10) [which picture of himself] doesn’t John know [which picture of himself] that Bill saw [which picture of himself]

Furthermore, binding ambiguities disappear whenever the wh-phrase is interpreted in situ for independent reasons. Example (11) is only ambiguous under the non-idiomatic reading, i.e., when take a picture is interpreted literally, meaning ‘pick up a picture’, instead of ‘photograph’.

(11) John wondered which picture of himself Bill took.

(12) a. **Non-idiomatic reading:**
John; wondered which picture of himself Bill; took.
b. **Idiomatic reading:**
John; wondered which picture of himself Bill; took.

Chomsky argues that whenever the wh-phrase forms a semantic unity with the verb, it must be interpreted in its base position. If this is correct, the idiomatic reading does not allow binding by John, as shown in (13b), because this requires that the wh-phrase be interpreted in its scope position. Under the non-idiomatic reading, on
the other hand, the *wh*-phrase can be interpreted in both positions (as in (13a)), and ambiguity is predicted.

(13)  a. *Non-idiomatic reading:*
       John wondered [which picture of himself] Bill took [which picture of himself]

       b. *Idiomatic reading:*
       John wondered [which picture of himself] Bill took [which picture of himself]

In sum, in order to be able to abolish the levels of D-Structure and S-Structure, Chomsky reformulates re-write rules in terms of Merge, and assumes that Binding Theory applies at LF and movement leaves copies visible at LF. In several respects, this even leads to an increase in empirical adequacy of the system.

1.2. Movement

In the preceding subsection, we have presented some of Chomsky's arguments in favor of the idea that Binding Theory need not refer to S-Structure. Let us now turn to the class of potential evidence in favor of S-Structure given in (4b), which concerns parametrization. Section 1.2.1 focuses on head movement. In section 1.2.2, movement of maximal projections will be examined.

1.2.1. Head movement

We argued in chapter 1, following traditional assumptions, that inflected verbs move to I (Aux) in French, whereas they stay in their base position inside the VP in English. This explains the distribution of VP adverbs such as *souvent* and *often*. The relevant structures are repeated in (14).

(14)  a. \([\text{IP} \text{Jean} [\text{VP often} [\text{VP Je} \text{embrasse}, \text{Marie}]]]\]

       b. \([\text{IP} \text{John} [\text{VP often} [\text{VP Je} \text{kisses Mary}]]]\]

In earlier stages of the theory, the verb used to be associated with inflectional morphemes by either raising V to I (French) or lowering I to V (English). These two types of affixation took place at S-Structure. However, the lowering operation in English goes against the uncontroversial assumption that moved elements must c-command their traces.

Chomsky (1995) circumvents this problem by adopting a lexicalist view on inflectional morphology whereby verbs enter the syntax fully inflected. He argues that the verb moves to I to license its inflection in an abstract fashion. This may happen overtly (French) or covertly (English). One way of implementing this idea is by assuming that the condition that requires V to move to I applies at S-Structure in French, and at LF in English. However, this would mean that we still need S-Structure.

Chomsky proposes a reformulation in terms of PF interpretability. He argues that the finite verb bares $e$-features and tense features which must be checked in the course of the derivation. Features are checked as soon as they are in a local relation
(see section 1.3 below) with a head endowed with matching features. Features matching those on the verb are located in I. Hence, the verb will have to move to I to have its features checked. This is rather similar to the OT analysis presented in chapter 1, according to which the inflected verb moves to I in order to be parsed in a position containing tense features.

According Chomsky, parametrization can be reduced to the fact that, depending on the language and the feature, features are checked either before or after Spell-Out. Chomsky distinguishes between strong and weak features. On the assumption that unchecked features as such are not interpretable at LF and unchecked strong features are not interpretable at PF, strong features must be checked before Spell-Out and weak features may be checked after Spell-Out.4

Thus, the difference between French and English in (14) can be reduced to feature strength. In French, the features on the verb are strong. As a consequence, the verb must move to I before Spell-Out. In English, on the other hand, these features are weak, which means that verb movement can take place covertly. Since, in any case, V-to-I movement must apply before LF (where both weak and strong features must have been checked), the examples in (14) have identical LF representations:

(15) Logical Form:
   a. \([IP\] Jean [I embrasse] [VP souvent [VP t_i Marie]]\]
   b. \([IP\] John [I kisses] [VP often [VP t_i Mary]]\]

By formulating the trigger for overt movement in terms of PF interpretability, Spell-Out remains an arbitrary point in the derivation: if it applies too soon, this will lead to a representation which is uninterpretable at PF and therefore ungrammatical. In other words, no reference is made to an internal interface such as S-Structure.

At this point, we are incorrectly predicting that English V-to-I raising is optional. Since the inflectional features are weak in this language, Spell-Out leads to an interpretable PF structure both before and after V has moved to I. However, verb movement in English does not apply overtly at all, which suggests that movement must apply after Spell-Out if possible. This follows from Chomsky’s principle of Procrastinate, based on the assumption that covert movement is cheaper than overt movement:

(16) Procrastinate (Chomsky 1995: 198):
Minimize overt syntax.

If overt movement is more expensive than covert movement, and if derivations must be as economic as possible (see section 2 below), then movement which can take place after Spell-Out must take place after Spell-Out. Thus, like in our OT analysis of chapter 1, parametrization of V-to-I movement is related to the interaction

---

4 Cf. Chomsky (1995: ch. 4), where it is argued that not all unchecked weak features are uninterpretable at LF. We leave this matter aside, since it is not directly relevant for the phenomena discussed here (see, however, chapter 3). See Broekhuis & Dekkers (to appear) for some speculations on feature interpretability and deletion at LF.
between economy and the requirement that V move to I (to check features or to satisfy a constraint).

At first sight, Procrastinate seems to refer to S-Structure in an implicit manner, since it makes an explicit distinction between overt and covert movement. However, in section 2.2, we will show that the intended effects of this principle can be maintained without the existence of S-Structure. Let us assume for now that (16) is indeed compatible with a grammar lacking this level of representation.

1.2.2. Movement of a maximal projection

If the approach sketched for parametrization of verb movement is correct, it can be extended to other word order phenomena. Let us take the examples in (17), and focus on the assignment of nominative Case and the position of the subject.

(17) a. French:
Marie voyait l'accident.
Marie saw the-accident
‘Marie saw the accident.’
b. Welsh:
Gwellodd Mair y ddamwain. (Ouhalla, 1991)
saw Mair the accident
‘Mair saw the accident.’

Suppose that these examples have the structures in (18), in which the French subject is in SpecIP and its Welsh counterpart in its base-position SpecVP (see Ouhalla, 1991). In both languages, the verb has moved to I. Since nominative Case is generally associated with Tense, let us further assume that I assigns Case to the subject in both languages. In a model of syntax that disposes of S-Structure, the structural relation that should hold between the subject and I at this level of representation can be parametrized: I assigns nominative Case to its specifier in French (Case assignment under spec-head agreement), and to a constituent in its c-command domain in Welsh (Case assignment under government). Alternatively, it could be that Case is uniformly assigned under spec-head agreement and that parametrization is due to the fact that this happens at S-Structure in French, and at LF in Welsh. LF representations are given in (19). The latter option again presupposes a lexicalist view on inflection.

(18) S-Structure/Spell-Out:
a. [IP Marie, [i voyait] [VP t, t l’accident]]
b. [IP [i gwellodd] [VP Mair t, y ddamwain]]

(19) Logical Form:
a. [IP Marie, [i voyait] [VP t, t l’accident]]
b. [IP Mair, [i gwellodd] [VP t, t y ddamwain]]

As a second alternative, we might generalize feature checking, and assume that nominative Case is licensed in the local domain of I (more particularly in SpecIP, see section 1.3 below) in all languages. The parametrization illustrated in (17)
would then be due to the fact that in some languages, the nominative Case feature is strong (French), while in others it is weak (Welsh). This leads to identical LF-structures (given in (19)), but distinct structures at Spell-Out (given in (18)). Again, parametrization is reduced to PF interpretability.

In short, generalizing the feature checking approach to movement of maximal projections leads to a uniform view on movement, in which a theory of Case assignment which refers to S-Structure is no longer needed.

1.3. Local relations

Feature checking always proceeds in the following way: some element $\alpha$, endowed with unchecked features, moves to the local domain of some head $X$ which bares matching features. The local domains of $X$ which Chomsky considers to be crucial in syntax are given in (20).\(^5\)

\[
\begin{align*}
\text{(20)} & \quad \text{a. Internal domain:} & \alpha \text{ is in the internal domain of } X & \text{iff } \alpha \text{ is the sister of } X. \\
& \quad \text{b. Minimal domain:} & \alpha \text{ is in the minimal domain of } X & \text{iff } \alpha \text{ is immediately dominated by a projection of } X. \\
& \quad \text{c. Minimal residue:} & \alpha \text{ is in the minimal residue of } X & \text{iff } \alpha \text{ is in the minimal domain of } X \text{ and } \alpha \text{ is not in the internal domain of } X.
\end{align*}
\]

Let us illustrate these definitions with the aid of the tree diagram given in (21). ZP is in the internal domain of $X$ because ZP is a sister of $X$. ZP is also in the minimal domain of $X$ because it is immediately dominated by $X'$, a projection of $X$ (notice that all elements in the internal domain of $X$ are also in the minimal domain of $X$). Consequently, it is not in the minimal residue of $X$. Since $W$ and YP are not sisters of $X$, these elements are excluded from the internal domain of $X$. The fact that they are, on the other hand, in the minimal domain of $X$ implies that they belong to the minimal residue of $X$.

\[
\begin{tikzpicture}
  \node {XP} child {node {YP} child {node {$W$}} child {node {$X$}} child {node {ZP}} child {node {$X'$}}};
\end{tikzpicture}
\]

\(^5\) The formulation of the definitions is ours. Domination should be understood here as an irreflexive relation that holds for categories rather than segments. Hence, in (21) below, $W$ is immediately dominated by $X'$, rather than by $X$ or itself. $\alpha$ and $\beta$ are sisters iff $\alpha$ and $\beta$ are immediately dominated by one and the same category. Note that these definitions only hold for trivial chains. See Chomsky (1995: ch. 3) for definitions of local domains of non-trivial head-movement chains.
(22) a. Internal domain of $X$: $\{ZP\}$  
b. Minimal domain of $X$: $\{W, YP, ZP\}$  
c. Checking domain of $X$: $\{W, YP\}$

Chomsky assumes that features on $\alpha$ are checked if $\alpha$ appears in the minimal residue of a head $X$ containing matching features. If $\alpha$ is a head, it is adjoined to $X$ (see section 1.2.1 above). If, on the other hand, $\alpha$ is a maximal projection, it should appear as the specifier of $X$ (see section 1.2.2 above). The internal domain is relevant for selection.

However, Bobaljik & Thráinsson (1997) argue that the checking domain should be identified with the minimal domain, rather than the minimal residue. From a minimalist perspective, this is superior to Chomsky's point of view, since it reduces the number of domains $C_{HL}$ refers to. If the internal domain is included in the checking domain, the notion minimal residue has lost its rationale and can be eliminated from the theory. Therefore, we will assume, until proof to the contrary is given, that only the internal and the minimal domain are local domains. Concretely, this means that, for instance, an internal argument can directly check the Case features of a verb for whom it is a sister.

If, indeed, syntactic relations are as local as suggested here, a notion such as head-government cannot be maintained, unless it is defined in terms of sisterhood. Among the P&P notions formulated in terms of government, Case Theory and the Empty Category Principle (ECP) figure most prominently. We have seen in section 1.2.2 that Case Theory can be reduced to Checking Theory. In chapter 4, we will examine the ECP (which crucially depends on a more liberal view on possible syntactic relations than the ones given above) in detail, and propose a reformulation in terms of economy.

2. Economy

Throughout Chomsky (1995: ch. 3-4), it is assumed (i) that movement is expensive, and that (ii) movement is more expensive before Spell-Out (Procrastinate, see section 1.2.1 above). However, it is difficult to determine how exactly Chomsky implements these general ideas in his computational system. This is the result of a conflict between two forces: on the one hand, it seems natural to interpret economy conditions in a global fashion, while on the other, Chomsky aims at a syntactic system that only disposes of local conditions.

Section 2.1 focuses on these two interpretations of economy of movement. In section 2.2, we will examine Procrastinate in more detail, and argue for a generalized and global approach to economy.

2.1. Global versus Local Economy

Economy conditions, which come in different flavors, can all be reduced to one general maxim: do not do anything costly. Applied to the operation Move, economy prevents movement. Since movement is a well-established syntactic transformation, economy cannot be absolute or inviolable; it is violated if this serves a higher purpose.
In Chomsky's computational system, this higher purpose is either the elimination of syntactic features as such or the construction of a representation that is interpretable at the interfaces. At first sight, these two goals seem interchangeable. At closer scrutiny, however, they are not. The choice between them depends on whether we choose for a global or a local interpretation of economy:

(23)  

a.  **Global Economy** (see Chomsky, 1995: 200):  
Movement is legitimate if it is necessary for convergence at LF.

b. **Local Economy** (Chomsky, 1995: 257)  
$\alpha$ can target $K$ only if a feature of either $\alpha$ or $K$ is checked by the operation.

If economy is interpreted globally, it is possible only after the derivation has been completed to evaluate whether all steps have been necessary for convergence. This presupposes so-called reference sets, which are comparable to candidates sets in OT. According to Chomsky (1995: 227), reference sets contain all possible convergent representations based on a single numeration (a set of lexical items, see section 1.1 above). Numerations are reminiscent of inputs in OT (see chapter 1). A simple way to ensure that economy is respected is by choosing the member of the reference set containing the smallest number of movement traces. It is crucial that reference sets only contain convergent derivations. Otherwise, representations lacking any movement traces would always be selected. To put this in OT terminology: convergence outranks economy.

Local Economy, on the other hand, applies at each step in the derivation. It requires that each individual movement transformation result in feature checking. Chomsky (1995: 228) argues that a local interpretation of economy reduces the computational complexity of the syntactic system because it would lead to a reduction of the number of derivations to be evaluated at the interface. Unfortunately, he does not prove this. Even if he had provided proof, Global Economy does not lead to a computational problem as long as we are able "to find a ready algorithm to reduce computational complexity", as Chomsky (1995: 228) remarks in relation to the fact that he is forced to adopt a global interpretation of Procrastinate (see section 1.2 above and section 2.2 below). We follow him on this point, and assume that as long as the increase in computational complexity induced by the global conditions of our grammar may be annulled by performance algorithms, computational matters are not relevant in a theory of linguistic competence. Another argument in favor of Local Economy could be that it does not presuppose the existence of numerations and reference sets. However, numerations and reference sets are required anyway if Procrastinate is a global condition.⁶

⁶ Note that Chomsky (1995: 227) assumes that Local Economy needs also a reference set and a numeration. This is related to the fact that he formulates the operation Move in terms of attraction and closeness. A target attracts the closest element that can check its features. Hence, at each point in the derivation, a reference set listing possible next steps is needed to determine which of them is most economical. Thus, we would end up with two types of reference sets: one for local evaluation (Local Economy), and one for global evaluation (Procrastinate). This is a considerable complication of the system. In this book, we will not go into the issue of the closeness condition on movement. Chomsky's formulation of this condition crucially relies on the assumption that movement always results in feature checking, which we will reject. See Ackema & Neeleman (1998) for an alternative, global interpretation of the closeness condition.
It would be incorrect to assume that Local Economy anticipates the uninterpretability of unchecked features at LF; it only reflects the idea that movement is feature-driven. An example will illustrate this point. Let us suppose that (24b) is the numeration that is needed to construct the sentence in (24a). Since (24b) contains all the right elements, (24a) converges. A minimally distinct numeration is given in (25b). This numeration does not lead to an interpretable LF (and PF) representation, since it does not contain an element that can check the nominative feature on T. In the optimal case, it will lead to (25a), which does not converge because it lacks a subject.

(24) a. Ethel kisses Billy.
   b. \{Billy_{[acc]}, kisses, Ethel_{[nom]}\}

   b. \{Billy_{[acc]}, kisses\}

Clearly, the ungrammaticality of (25a) is the result of LF (and PF) uninterpretability, rather than of Local Economy. This shows that Local Economy forces us to adopt two independent conditions which both refer to feature checking: (i) a principle stating that representations containing unchecked features do not converge, and (ii) Local Economy itself. In this respect, Global Economy is the more general principle, since it does not refer to feature checking. Besides the fact that general nature of Global Economy is attractive in itself, it has two other advantages. First, it allows us to have economy interact with more than only the requirement that features be checked, which increases the empirical scope of the system, as we will see in section 3 as well as in subsequent chapters. Second, Global Economy subsumes Procrastinate. Let us start with the latter point.

2.2. Generalized Economy

In section 1.2, we have seen that Chomsky attributes parametrization to the interaction between the economy condition Procrastinate, repeated here in (26), and feature strength.

(26) Procrastinate (Chomsky, 1995: 198):
    Minimize overt syntax.

Chomsky considers Procrastinate a global principle (see in particular Chomsky, 1995: 228). This is not so surprising, since if it were local, it would have to be formulated along the lines of (27), which is modeled on (23b). Clearly, the local (27) is unacceptable. It refers to Spell-Out, which amounts to an implicit re-introduction of the level of S-Structure.7

---

7 In Chomsky (1995: ch. 4), it is assumed that strong features are not located on the moved element, but rather on the attracting head (see footnote 6). Overt subject movement, for instance, is due to a strong feature on I. Chomsky (1995: 234) notes that “the descriptive property of strength is \{(i)\}”.

(i) D is canceled if α [endowed with a strong feature] is in a category not headed by α.
Local Procrastinate:
α targets K before Spell-Out only if a strong feature of either α or K is checked by the operation.

This also holds for the original definition of Procrastinate in (26), which refers to overt syntax. In this light, we should wonder whether we really need Procrastinate as an independent principle. Notice that (26) can be reformulated as in (28b). This formulation is similar to that of Global Economy in (23a), repeated here as (28a). It would be interesting to see if the two conditions in (28) can be reduced to one single economy principle.

(28) a. Global Economy:
Movement is legitimate if it is necessary for convergence at LF.

b. Procrastinate:
Overt movement is legitimate if it is necessary for convergence at PF.

The simplest way to implement this idea is by taking LF representations as inputs for the PF component. This allows us to have both conditions in (28) operate on one and the same reference set, which is the first step toward a single economy condition. In fact, this can be considered a natural extension of Chomsky’s (1995: ch. 4) views on movement. He argues that the operation Move is primarily interested in formal features. Therefore, in the minimal case, Move only affects the formal features that should be checked. Case feature checking, for instance, would only involve movement of the Case feature. Overt movement departs from this minimal scenario, since it involves movement of entire categories (consisting of syntactic structure, as well as categorial, phonological, and, possibly, semantic features). Clearly, category movement is more expensive than movement of isolated formal features. If only category movement can check strong features, we can reformulate Global Economy and Procrastinate in terms of feature and category movement:8

(29) a. Global Economy (reformulated):
Formal features move if this is necessary for convergence at LF.

b. Procrastinate (reformulated):
Formal features carry along their category if this is necessary for convergence at PF.

These two conditions in (29) can be reduced to a single one, based on the idea that movement of syntactic material in general is expensive:

The statement in (i) expresses that the maximal projection of a head X can only be embedded in a larger structure if the strong features of X have been checked. If elements can only be embedded in a larger structure before Spell-Out (which follows if all members of the numeration must have been used before Spell-Out), (i) ensures that strong features must be checked before Spell-Out, except when (and this is remarkable) they are located in the highest head of the total structure. Be this as it may, (i) should not be considered a principle that can be substituted for Procrastinate, since it does not follow from (i) that covert movement must apply after Spell-Out.

Chomsky (1995: 262) postulates the principle in (i), which has the same effect as (29b).

(i)  F [a formal feature] carries along just enough material for convergence.
According to (30), no movement at all is preferred. If a formal feature moves, or even worse, if a formal feature moves and carries along its category, this must lead to convergence.

Now that we reduced the two economy conditions to one general one, we need only one syntactic structure for both interfaces. Instead of making a distinction between overt and covert movement, we distinguish between movement of formal features in isolation and movement of formal features carrying along the category. Let us return to our French and Welsh examples in (17), repeated here as (31).

(31) a. Marie voyait l'accident.
    Marie saw the-accident
b. Gwellodd Mair y ddamwain.
    Saw Mair the accident

As stated in section 1.2.2, the subject moves to SpecIP in French, while it remains in SpecVP in Welsh. Although the Welsh subject is not parsed in the checking domain of I, it is associated with it for Case and agreement purposes. To put this in terms of feature movement: the Case and q-features of the Welsh subject move to the checking domain of I in isolation. Chomsky (1995: 270) argues that because isolated features do not qualify as maximal projections, they behave like heads and adjoin to the target. In French, the formal features of the subject pied-pipe the entire category. Thus, the examples in (31) have the structures in (32).

(32) a. \[IP Marie, [1 voyait] [t_f l'accident]\]
    b. \[IP [FF(Mair) [gwellodd]] [t_f ddamwain]\]

Let us take a closer look at the evaluation procedure. Recall that Chomsky assumes that economy conditions evaluate reference sets, and that a reference set contains all structures and only those structures built from an identical set of lexical items (a numeration). The numerations underlying the structures in (32) are given in (33).

(33) a. N = \{voyait, l', Marie, accident, F*\}
    b. N = \{gwellodd, y, Mair, ddamwain, F*\}

A subset of representations that can be constructed from these two numerations is given in (34). Some of these structures are excluded from the reference set. In (34a-a'), the features of the subject are in situ. These representations contain unchecked features. As a consequence, they cannot be interpreted at LF. Representations which are uninterpretable at LF are marked with the symbol \#.

\[FF_1(a)\] stands for (a subset of) the formal features of a, the chain (\(FF_1(a)_t\), \(a_t\)) is the result of movement of (a subset of) formal features of a, and the chain (\(a_i, t_j\)) is brought about by movement of the category of a.

\[F^*\] represents the functional heads in N. See section 4.2 below for a discussion of related issues.
category movement of the subject although this movement is necessary because the nominative Case feature is strong in English. Uninterpretability at PF is marked with the symbol $\hat{\phi}$. Reference sets only contain representations which are not marked with $\phi$ or $\hat{\phi}$; the Welsh reference set contains options (b'), (c'), (d'), (e'), whereas (c), (d), (e) are members of the French reference set.

(34) No movement of (the formal features of) the subject
a. $\hat{\phi} \phi [\text{IP} [I_1 \text{voyait}] [\text{VP} \text{Marie} t_1 \text{l''accident}]]$

Movement of the formal features of the subject:

b. $\hat{\phi} [\text{IP} [I_1 \text{FF(Marie)} t_1 \text{voyait}] [\text{VP} \text{Marie} t_1 \text{l''accident}]]$

c. $[\text{IP} \text{Marie} [I_1 \text{voyait}] [\text{VP} t_1 t_2 \text{l''accident}]]$

c'. $[\text{IP} \text{Marie} [I_1 \text{gwellodd}] [\text{VP} t_1 t_2 \text{ddamwain}]]$

Movement of the entire subject and movement of the formal features of the object:

d. $[\text{IP} \text{Marie} [I_1 \text{FF(l'accident)} t_1 \text{voyait}] [\text{VP} t_1 t_2 \text{l''accident}]]$

d'. $[\text{IP} \text{Marie} [I_1 \text{FF(y ddamwain)} t_1 \text{gwellodd}] [\text{VP} t_1 t_2 \text{y ddamwain}]]$

Movement of the entire object and movement of the formal features of the subject:

e. $[\text{IP} \text{l'accident} t_1 \text{FF(Marie)} t_1 \text{voyait}] [\text{VP} \text{Marie} t_2 t_3 \text{l'accelent}]]$

e'. $[\text{IP} \text{y ddamwain} t_1 \text{FF(Marie)} t_1 \text{gwellodd}] [\text{VP} \text{Marie} t_2 t_3]]$

The options in (34) are given in order of costliness. In accordance with Generalized Global Economy, we should simply pick the highest-ranked representation not marked with the symbols $\phi$ or $\hat{\phi}$, i.e. the French (34c) and the Welsh (34b'). Notice that movement that does not result in feature checking is illegitimate since there will always be cheaper representations available in the reference set. Recall that we proposed in section 1.3 that heads can check the features of their complements. If this is correct, accusative case can be checked in situ. As a result, the structures in (34d-e') involve unnecessary movement of (the formal features of) the object. Consequently, these structures are excluded by (30).

3. Toward constraint interaction

This section focuses on the possibility of reducing movement parametrization to constraint interaction, rather than feature strength. This leads to a more natural view on movement parametrization (section 3.1) as well as to an extension of the empirical scope of the system (section 3.2).

3.1. Parsing phonological features

Thus far, we have assumed that pied-piping contributes to PF interpretability because isolated feature movement is insufficient to check strong features and unchecked features are uninterpretable at PF. However, it is not clear why isolated
feature movement does not lead to strong feature checking and why strong features are uninterpretable at PF.

It is possible to establish a more direct relation between pied-piping and PF interpretability. If the position of formal features is first and foremost relevant at LF, and if the position of phonological features is a PF matter, pied-piping ensures that LF information is pronounced at PF. This suggests that the form of the syntactic surface structure is determined by the two forces given in (35). According to (35a), (phonological) features must stay where they are, whereas (35b) forces them to move wherever their formal features move.

(35)  

a. STAY: Do not move syntactic material.  
b. PARSE-F: Pronounce constituents in the same position as their formal feature F.

In OT, condition (35a) can be interpreted as a violable constraint (see chapter 1), and (35b) as a family of violable constraints. Each member of this family (PARSE-wh, PARSE-Case, etc.) requires that phonological features be parsed in the same position as the individual formal feature. OT-style constraint rankings can now be held responsible for movement parametrization. If the individual parse constraint outranks STAY, the formal feature in question pied-pipes the category; if STAY outranks the parse constraint, the formal feature moves in isolation. Applying this to the example above, PARSE-Case outranks STAY in French, while the reverse is true for Welsh:

(36)  

a. French: PARSE-Case » STAY  
b. Welsh: STAY » PARSE-Case

Let us examine the evaluation of the French and Welsh examples in (31) in more detail. Recall (see section 2.1) that a candidate/reference set contains all structures produced by Gen/C on the basis of a given numeration which converge at LF, and only those structures (this assumption will be revised in section 4.2). The evaluation of the French example is given in (37). In this tableau, a non-exhaustive list of output analyses of the input is given. Movement of formal features incurs one violation of STAY, and pied-piping incurs two. PARSE-Case is violated whenever the subject is not parsed with its formal features in the checking domain of I. We have established that in French, PARSE-Case outranks STAY. As a result of this ranking, candidates (37a) and (37d) will lose because they violate PARSE-Case. STAY has to decide between the remaining two candidates. Since (37c) incurs one more mark on STAY than (37b) does, the latter is optimal. Notice that
this additional violation of STAY is caused by the fact that candidate (37c) involves movement of the formal features of the object to I, whereas candidate (37b) does not.

(37)  

<table>
<thead>
<tr>
<th>French (N = {voyait, l', Marie, accident, F*})</th>
<th>P-Case</th>
<th>STAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ([I_p [I FF(Marie), [I voyait,]] [VP Marie, t_i l'accident]])</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>b. ([I_p Marie, [I voyait,]] [VP t_i t_k l'accident]])</td>
<td>**</td>
<td>—</td>
</tr>
<tr>
<td>c. ([I_p Marie, [I FF(l'accident), [I voyait,]] [VP t_i t_k l'accident]])</td>
<td>***!</td>
<td>—</td>
</tr>
<tr>
<td>d. ([I_p l'accident, [I FF(Marie), [I voyait,]] [VP Marie, t_k t_i]])</td>
<td>*!</td>
<td>***</td>
</tr>
</tbody>
</table>

Welsh is characterized by the opposite ranking. This leads to the evaluation in (38). In this tableau, it is STAY that plays a decisive role. Candidate (38a) is optimal because it violates STAY only once, whereas the other candidates do so twice or more.

(38)  

<table>
<thead>
<tr>
<th>Welsh (N = {gwellodd, y, Mair, ddamwain, F*})</th>
<th>STAY</th>
<th>P-Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ([I_p [I FF(Mair), [I gwellodd,]] [VP Mair, t_i y ddamwain]])</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>b. ([I_p Mair, [I gwellodd,]] [VP t_i t_k ddamwain]])</td>
<td>**!</td>
<td>—</td>
</tr>
<tr>
<td>c. ([I_p Mair, [I FF(y ddamwain), [I gwellodd,]] [VP t_i t_k y ddamwain]])</td>
<td>*<em>!</em></td>
<td>—</td>
</tr>
<tr>
<td>d. ([I_p y ddamwain, [I FF(Mair), [I gwellodd,]] [VP Mair, t_k t_i]])</td>
<td>*<em>!</em></td>
<td>*</td>
</tr>
</tbody>
</table>

Two remarks are in order. First, tableaux (37) and (38) clearly illustrate that STAY actively constrains both movement of formal features and pied-piping of phonological features. In (37), STAY decides between candidates (b) and (c), since it disprefers movement of formal features. In (38), on the other hand, STAY distinguishes between movement of formal features in isolation and pied-piping of phonological features, with the result that candidate (a) rather than candidate (b) is optimal.

Second, candidates involving movement of formal features in isolation which does not result in feature checking are harmonically bound by candidates which do not involve such movement (see chapter 1 for a definition of harmonic boundedness). Harmonically bound candidates lose under any constraint ranking. Hence, movement of formal features always results in feature checking.

3.2. Feature strength versus constraint interaction

From an empirical point of view, the choice between feature strength or constraint ranking is innocuous as long as we restrict ourselves to parametrization of feature-
driven movement. However, there are other syntactic forces that determine the input to the PF component. These concern linearity, deletion, insertion, silent structure, superfluous structure, information packaging, and so on. By way of an illustration, a preliminary discussion of some of the issues that will be examined in detail in subsequent chapters are presented in this section.

Section 3.2.1 focuses on deletion in the complementizer domain, one of the central issues of this book. Whereas deletion can successfully be reduced to constraint interaction, as Pesetsky (1997, 1998) shows, it is doubtful if this phenomenon can be captured in terms of feature strength. In section 3.2.2, we will study the influence of information structure on word order, and conclude that not all aspects of movement can be stated in terms of feature checking. We will return this issue in chapter 5.

### 3.2.1. The complementizer domain

Pesetsky (1997, 1998) observes that linear order plays a central role in complementizer deletion. Consider the French examples in (39). In the absence of an overtly realized pronoun, the complementizer *que* is pronounced, as in (39a). Example (39b) illustrates, on the other hand, that whenever the relative pronoun is pronounced, *que* is absent.

\[(39)\]
\begin{align*}
\text{a. } & \text{ la femme que je connais} \\
& \text{the woman that I know} \\
& \text{‘the woman I know’} \\
\text{b. } & \text{ la femme avec qui j’ai dansé} \\
& \text{the woman with whom I have danced} \\
& \text{‘the woman I danced with’}
\end{align*}

Pesetsky argues that (externally headed) relative clauses should be uniformly associated with the structure in (40). If so, the examples in (39) correspond to the structures in (41). In (41a), the relative pronoun is deleted, while the complementizer is pronounced. (41b) shows the reverse deletion pattern.

\[(40)\quad \left[ \text{CP} \left( \left[ \text{P} \text{ pronoun} \right] \left[ \text{C} \text{ complementizer} \right] \right) \right] \]

\[(41)\]
\begin{align*}
\text{a. } & \left[ \text{CP} \left[ \text{qui} \right] \left[ \text{C} \text{ que} \right] \right] \left[ \text{IP} \ldots \right]^{14} \\
\text{b. } & \left[ \text{CP} \left[ \text{avec qui} \right] \left[ \text{C} \text{ que} \right] \right] \left[ \text{IP} \ldots \right]
\end{align*}

Traditionally, these facts are analyzed in terms of filters. Chomsky & Lasnik (1977, henceforth C&L) argue that clauses in which both the relative pronoun and the complementizer are pronounced violate the Doubly Filled Comp Filter (DFCF) given in (42). The DFCF presupposes an exocentric approach to phrase structure, in which both operators and complementizers occupy the Comp position. The DFCF

---

13 Chomsky (1995: ch.4, fn. 66) suggests that phenomena like complementizer pronunciation and verb-second follow from phonological considerations, rather than from feature checking. This is in line with the spirit of this book.

14 See Kayne (1976) for arguments in favor of analyzing *que* as a complementizer, rather than a relative pronoun. Notice that the deleted relative pronoun in SpecCP is identified with *qui*, which will be motivated in chapter 3.
states that each Comp may contain at most one element. If the examples in (39) have the underlying structure in (42b), then the DFCF requires that either *avec qui or *que be deleted.

(42) a. *[Comp w$\cdot$-phrase $\varphi$], $\varphi \neq e$
    b. $[S\ Comp (avec) qui que] ...$

In chapter 3, we will go into the empirical flaws of the DFCF (or any equivalent filter formulated in terms of endocentric phrase structure). For now, it suffices to note that according C&L, these and similar instances of deletion call for an output-oriented analysis. They treat them as "contextual dependencies that cannot be formulated in the narrower framework of core grammar" (C&L: 432). The notion core grammar stands for the transformational component in C&L's framework, which corresponds to Chomsky's (1995) C$_{HL}$ or Gen in OT. If C&L's statement is correct, an analysis in terms of output constraints is promising.

According to Pesetsky, patterns like the one in (41) follow from the interaction of the three constraints in (43). REC makes a distinction between deletion of relative pronouns embedded in a PP and deletion of those that are not. Pesetsky assumes that prepositions have semantic content, whereas relative pronouns do not (see also C&L: 466). As a result, the deletion of PPs containing a relative pronoun violates REC, whereas that of bare relative pronouns does not. The second constraint, LE(CP), is an alignment constraint (see McCarthy & Prince, 1993) which is violated whenever the first pronounced element in a CP is not a complementizer or some other function word from the extended projection of the verb. Finally, TEL prohibits the pronunciation of complementizers and other function words.

(43) a. RECOVERABILITY (REC): A syntactic unit with semantic content must be pronounced.\(^{15}\)
    b. LEFTEDGE(CP) (LE(CP)): The first pronounced word in CP is a function word related to the main verb of the CP.
    c. TELEGRAPH (TEL): Do not pronounce function words.

Pesetsky argues for French that these constraints are ranked as in (44).

(44) French: REC $\gg$ LE(CP) $\gg$ TEL

\(^{15}\) We simplified Pesetsky's definition of REC. The original definition, given in (i), is too complex to qualify as a possible OT constraint, since it expresses matters that can be captured in terms of interaction of simpler constraints (cf. Broekhuis & Dekkers, to appear).

Furthermore, the definition in (i), because it allows deletion of PPs in the context of a local antecedent, incorrectly predicts that (iiib) should be acceptable next to (iia) (see chapter 3 for an analysis of English relative clauses). The formulation in (43), on the other hand, correctly excludes (iiib), as pointed out by João Costa (p.c.).

(ii) a. John cuts the meat with the knife with which he has cut the cheese.
    b. *John cuts the meat with the knife (that) he has cut the cheese.
Let us add the operation \textit{Delete}, given in (45), to Gen/C_{HL}. This operation can freely be applied during the derivation.

\begin{equation}
\text{Delete: remove the phonological matrix of syntactic units.}
\end{equation}

Restricting ourselves to the complementizer domain, we now have the choice between the four options in (46).

\begin{equation}
\begin{align*}
\text{a. } & \text{[CP [P pronoun] [c complementizer] [IP ...]]} \\
\text{b. } & \text{[CP [P pronoun] [c complementizer] [IP ...]]} \\
\text{c. } & \text{[CP [P pronoun] [c complementizer] [IP ...]]} \\
\text{d. } & \text{[CP [P pronoun] [c complementizer] [IP ...]]}
\end{align*}
\end{equation}

Evaluations are given in tableaux (47) and (48). In tableau (47), deletion of the relative pronoun does not violate REC because it is not embedded in a PP. Furthermore, \textit{LE(CP)} prefers elements in SpecCP to be deleted because this ensures that the complementizer occurs in CP-initial position. As a result, candidate (47c) is optimal, and \textit{TEL} is irrelevant.

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|}
\hline
\textit{French} & \text{REC} & \text{LE(CP)} & \text{TEL} \\
\hline
a. la femme \text{[CP qui [c que] [IP je connais]]} & * & * & * \\
b. la femme \text{[CP qui [c que] [IP je connais]]} & * & * & * \\
c. la femme \text{[CP qui [c que] [IP je connais]]} & * & * & * \\
d. la femme \text{[CP qui [c que] [IP je connais]]} & * & * & * \\
\hline
\end{tabular}
\caption{Tableau (47) Evaluations}
\end{table}

In tableau (48), a different picture emerges. When \textit{avec qui} is deleted, the preposition is not recoverable, which excludes the candidates (48c) and (48d). The second constraint, \textit{LE(CP)}, does not decide between the two remaining candidates (48a) and (48b) because it is violated by both. In this case, \textit{TEL} plays a crucial role because it prefers candidate (48b) over candidate (48a): only the former involves the deletion of the function word \textit{que}. This is an example of \textit{the emergence of the unmarked} (see McCarthy & Prince, 1994): due to the low rank of \textit{TEL}, the unmarked status of structures that do not contain function words can be witnessed on the surface only in a subset of syntactic contexts.

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|}
\hline
\textit{French} & \text{REC} & \text{LE(CP)} & \text{TEL} \\
\hline
a. la femme \text{[CP avec qui [c que] [IP j'ai dansé]]} & * & * & ! \\
b. la femme \text{[CP avec qui [c que] [IP j'ai dansé]]} & * & * & ! \\
c. la femme \text{[CP avec qui [c que] [IP j'ai dansé]]} & * & * & ! \\
d. la femme \text{[CP avec qui [c que] [IP j'ai dansé]]} & * & * & ! \\
\hline
\end{tabular}
\caption{Tableau (48) Evaluations}
\end{table}

In subsequent chapters, it will be shown that these three constraints explains a range of related phenomena, and that their mutual ranking is responsible for
differences between languages. Now, if feature strength can only account for movement parametrization, while constraint interaction can also account for deletion, the latter should be preferred to the former.

3.2.2. Information packaging

Another domain of syntax which is hard to capture in terms of feature checking is information packaging, i.e. the syntactic encoding of pragmatic information. It has often been pointed out that in cases like (49) (Heavy NP Shift) and (50) (Italian “Free” Inversion), word order optionality is only apparent. The application of these “rules” is not optional if their interpretational effects are taken into account. Although C&L do not discuss these cases, they would probably analyze them in terms of stylistic rules (see Rochemont, 1978), which, like deletion and filters, are located outside the transformational component.

(49) Zubizarreta (1994):
   a. Max put all the boxes of home furnishings in his car.
   b. Max put in his car all the boxes of home furnishings.

   a. Gianni ha gridato.
      Gianni has screamed
      ‘Gianni has screamed.’
   b. Ha gridato Gianni.

Let us start with the examples in (49). Example (49a) displays the unmarked order. In (49b), all the boxes of home furnishings is interpreted as narrowly focused.\(^{16}\) This pragmatic information is marked syntactically by the inverted order of the NP and the PP. This inversion can be analyzed in two ways. Either the NP is moved rightward, or the PP is moved leftward. In either case, a syntax based on the hypothesis that movement applies in order to check features (Chomsky, 1995) will not be able to map this pragmatic information onto syntactic structure in a straightforward manner. Let us suppose that focused elements are endowed with a [+foc] feature and, therefore, must move to the checking domain of a Focus Phrase. Now, if the NP is moved, it should end up in a right-branching specifier of this FocP to check its feature. However, right-branching specifiers are not uncontroversial from a conceptual point of view. If we allow them in our theory, they are not likely to occur in an SVO language like English. If the PP moves, on the other hand, it is even less likely that the NP will be able to check its [+foc] feature.

It seems more plausible that the focus interpretation of the NP in (49b) is due to its right-peripheral position. In OT syntax, a significant amount of attention has been paid to the crosslinguistic tendency to align focused constituents with the

---

\(^{16}\) By ignoring the influence of prosody (the weight of the shifted NP, the role of stress assignment, etc.) on word order, we are simplifying matters considerably. Possibly, prosody mediates between information structure and syntax. For instance, focus has an affinity with primary stress (on utterance level) and primary stress with the right-periphery of the sentence (see Zubizarreta, 1994, among others).
right edge of the clause, which is generally attributed to an alignment constraint such as ALIGN-foc given in (51a) (see Broekhuis, 1998; Costa, 1996, 1997, 1998; Dekkers, 1997; Grimshaw & Samek-Lodovici, 1995; Samek-Lodovici, 1996). This constraint is in conflict with the economy constraint STAY. STAY can be held responsible for the fact that (49b) is syntactically marked. Irrespective of whether it is the NP or the PP that moves, (49b) involves one more movement step than (49a). The fact that this order surfaces indicates that ALIGN-foc outranks STAY.

\[(51)\]

\[\begin{align*}
\text{a. ALIGN-foc:} & \quad \text{Align focused constituents with the right edge of the verbal extended projection.} \\
\text{b. STAY:} & \quad \text{Traces are prohibited.}
\end{align*}\]

Like in (49), the difference in word order in (50) gives rise to interpretational differences. In (50b), the subject is focused, whereas in (50a) it is not. Again, focus interpretation is associated with the right edge of the sentence. There seem to be two forces at work. On the one hand, focused constituents tend to appear in clause-final position to satisfy ALIGN-foc.\(^\text{17}\) On the other hand, PARSE-Case requires that subjects be parsed in SpecIP (section 3.1 above).

In Italian, ALIGN-foc outranks PARSE-Case. The examples in (50) show that focused subjects appear at the right edge of the sentence. In English, on the other hand, PARSE-Case outranks ALIGN-foc. Irrespective of whether the subject is focused or not, it appears in preverbal position:

\[(52)\]

\[\begin{align*}
\text{a. *Has screamed John.} \\
\text{b. John has screamed.}
\end{align*}\]

The evaluations are given in tableaux (53) and (54).

\[(53)\]

\[
\begin{array}{|c|c|}
\hline
\text{Italian} & \text{A-foc} & \text{P-Case} \\
\hline
\text{a. |IP | F(Gianni) | ha] gridato Gianni_{[foc]} |} & \text{fa} & \text{*} \\
\text{b. [IP Gianni_{[foc]} | ha] gridato]} & \text{*!} \\
\hline
\end{array}
\]

\[(54)\]

\[
\begin{array}{|c|c|}
\hline
\text{English} & \text{P-Case} & \text{A-foc} \\
\hline
\text{a. |IP | F(John) | has] screamed John_{[foc]} |} & \text{*!} \\
\text{b. [IP John_{[foc]} | has] screamed]} & \text{fa} & \text{*} \\
\hline
\end{array}
\]

It seems problematic to reduce information packaging to feature strength. According to C&L, “stylistic rules” apply outside the scope of the transformational component. Nevertheless, these rules lead to differences in word order brought about

\(^{17}\) This analysis is largely inspired by Grimshaw & Samek-Lodovici (1995). They assume that the postverbal subject in Italian is adjoined to VP. We will not go into this matter, since satisfaction or violation of ALIGN-foc depends on the linear, rather than the hierarchical position of the focused element. Accordingly, the syntactic structure of the candidates in tableaux (53) and (54) is underspecified.
by movement transformations. If we choose to adopt the feature strength approach, we have to assume a second transformational component (post-CHL) responsible for phenomena such as the ones examined in this subsection.

However, depending on the language, information packaging can have a substantial influence on word order, and, for instance, annihilate the effect of Case-driven movement. This suggests that information-driven movement should not be treated as belonging to the periphery of language contrary to what the introduction of the second component implies. More importantly, post-CHL movement as such is a complication of the system which should be avoided if possible. In OT, on the other hand, the degree of discourse-configurationality follows from the ranking of discourse-configurational constraints such as ALIGN-foc. Since these directly interact with non-discourse-configurational constraints such as PARSE-Case, no second transformational component is needed.

In combination with the conclusions drawn in the previous subsection: if we substitute constraint ranking for feature strength, information packaging and deletion can be accounted for. This enlarges the scope of the system considerably and no auxiliary modules of syntax need to be postulated.

4. Derivations & Evaluations

We have constructed a model of syntax (henceforth Derivations & Evaluations or D&E) consisting of two parts: Gen/CHL is responsible for deriving representations, which are subsequently evaluated with respect to a constraint ranking. Both Chomsky’s system and “standard” OT syntax are set up in a similar way: syntactic objects are generated and subsequently evaluated.

D&E resembles “standard” OT syntax in the sense that the evaluation procedure is more powerful than Chomsky assumes. Constraints conflict with each other, and their ranking explains syntactic patterns within and across languages. However, Gen plays a more prominent role in D&E than it does in “standard” OT syntax. D&E acknowledges that any generative grammar must be explicit about the way in which structures are produced, even if the explanatory power of the grammar is not primarily associated with this generative device.

The properties attributed to OT syntax in chapter 1 also hold for D&E. In addition, Gen is identified with CHL. In this section, we will examine three issues that deserve further attention. Section 4.1 focuses on feature checking in Extended X-bar Theory, a point that is specific for D&E. Subsequently, we will address two topics that are relevant in OT syntax in general: candidate set membership (section 4.2) and the division of labor between Eval and Gen (section 4.3).

4.1. Features and Extended X-bar Theory

We have argued in chapter 1, following Grimshaw (1997), that functional structure is freely available in Gen, and that constraint ranking determines how much of this
structure is really needed in a language.\footnote{See Nash \& Rouveret (1996) for a non-OT proposal along the same lines. They introduce so-called proxy heads, empty heads which are available, if necessary, for feature checking purposes.} This is also the view adopted in D&E. Ideally, all syntactic information is associated with the lexical head. However, we depart from this ideal situation by assuming that in the verbal extended projection, there is one functional head that is intrinsically associated with features, i.e. I (see Broekhuis, 1998, for a similar point of view).

All other functional heads lack formal features. Nevertheless, they can project, if necessary, for the purpose of feature checking. This follows if the notion \textit{minimal domain} (see section 1.3 above) is defined in terms of extended projections rather than perfect projections, which leads to the definition in (55b) (cf. (20b)). The definition of the notion \textit{internal domain} remains unchanged.

\begin{enumerate}
    \item Internal domain: \( \alpha \) is in the internal domain of \( X \) iff \( \alpha \) is the sister of \( X \).
    \item Minimal domain: \( \alpha \) is in the minimal domain of \( X \) iff \( \alpha \) is immediately dominated by an extended projection of \( X \).
\end{enumerate}

In (56), ZP is immediately dominated by CP. It follows from (55b) that ZP is in the minimal domain of I, since CP is an extended projection of I. If, as we argued in section 1.3, the checking domain is identified with the minimal domain, ZP can check features on I.

\begin{figure}
\centering
\begin{tikzpicture}
    \node (CP) {CP}
    child {node (ZP) {ZP}
        child {node (C') {C'}
            child {node (C) {C}}
            child {node (IP) {IP}
                child {node (I) {I}}}}}
    child {node (IP) {IP}
        child {node (I) {I}}};
\end{tikzpicture}
\end{figure}

In this book, we will restrict ourselves to interrogative and relative features. Let us assume, following Rizzi (1991), that these are located in I. As such, they may be checked by an element in either SpecIP or SpecCP.\footnote{Along the same lines, features on V may be checked in a higher projection. In ECM-constructions, for instance, the object of the matrix verb is the external argument of an embedded predicate, and therefore not in the minimal domain of its Case-assigner. This may trigger movement to a specifier position in the functional domain of V corresponding to SpecAgrOP. Similarly, negation may turn out to involve the checking of a feature on V. We leave this for further research.} The actual position in which these features are checked depends on constraint ranking. For instance, the English ranking requires that interrogative subjects be parsed in SpecIP, and interrogative objects in SpecCP, as we saw in chapter 1. Similar conclusions are
drawn for relative clauses in chapter 3. In chapter 5, we will argue that also non-subjects may move to SpecIP to check features.

Whereas overt wh-movement can target both SpecIP and SpecCP, overt nominative Case movement always seems to be to SpecIP. Let us tentatively assume that operators such as wh-phrases cannot appear in SpecIP if the subject checks nominative Case in SpecCP because wh-phrases must c-command the whole clause that is interpreted as being in their scope. If this requirement is not met, the sentence will be uninterpretable in the semantic component.20

4.2. Candidate set membership

Up until now, we have been assuming that structures are in competition with each other if they are based on a single numeration (a set of words). However, Broekhuis & Dekkers (to appear) notice that the acceptability of both examples in (58) is incompatible with this assumption. These examples would have to be in one candidate set because they are based on the same numeration, which is given in (57).

(57) \[ N = \{ \text{game, Bob, the, wonders, thinks, who, Hank, lost} \} \]

(58) a. Bob wonders [who; Hank thinks [\(i\); lost the game]]

b. Bob thinks Hank wonders [who; lost the game]

However, it is not likely that the sentences in (58) compete with each other. If these sentences were in competition, (58a) would lose because it contains more movement traces than (58b). The fact that both structures in (58) do in fact surface suggests that they survive distinct evaluations. Hence, the assumption that members of candidate sets are based on identical numerations is too weak.

Notice that although the two sentences are composed of the same words, they receive distinct semantic interpretations. In this light, it seems plausible that the semantics of output structures play a crucial role in the evaluation procedure in one way or another. And indeed, there exists a broad consensus that semantic interpretation has an influence on the evaluation procedure, although the formal implementation of this idea is subject to debate. Two viewpoints are found in the OT literature.

Some authors (e.g. Ackema & Neeleman, to appear; Broekhuis & Dekkers, to appear) argue for a semantic identity requirement on candidate sets: only structures that have an identical (or non-distinct) semantic interpretation may enter in competition with each other. Thus, the two structures in (58) are in distinct candidate sets and will not be in competition, simply because they have different meanings. This approach presupposes an algorithm that prevents the comparison of

20 Traditionally, it is assumed that there are designated positions for each type of element that should be licensed (operators are licensed in SpecCP, subjects are licensed in SpecIP, etc.). In Chomsky (1995), this follows from the assumption that features are located in fixed positions (operator features are located in C, whereas the nominative Case feature is located in I, therefore operators check features in SpecCP, and subjects in SpecIP). At first sight, this rigid view on syntax seems superior because it excludes the possibility of licensing nominative elements in SpecCP. Be this as it may, it still leaves us with the question of why elements are licensed or checked in these fixed positions. Furthermore, it also forbids operators to be licensed in SpecIP. One of the objectives of this book is to show that this is an incorrect assumption.
semantically distinct structures. However, it is unclear how to construct such an algorithm. Moreover, it is not obvious that we really need it because only syntactically relevant aspects of meaning matter in candidate evaluation.

This is implicitly acknowledged by the proponents of the alternative approach, who assume that structures are evaluated with respect to a(n underspecified) semantic target specified in the input. According to Legendre et al. (1995: 610), “Inputs consist of […] skeletal structures containing predicate-argument structure and scope information”. Hence, the structures in (58) are optimal analyses of distinct inputs. The structure in (58a) is faithful to the predication and scope information given in (59a), and the one in (58b) to that in (59b).

(59)  
\[ \ldots \text{V} \ldots [Q_j \ldots \text{V} \ldots [x_i \text{V} \ldots]] \]

This presupposes a syntax for inputs: if the input is structured, this structure must be generated in one way or another. In other words, we would need a second generator, a significant complication of our system. However, Legendre et al. (1995: 610) state that “each output for an input contains that input” (the scopal information in (59a), for instance, is contained in the output in (58a)), which could mean that this complication is unnecessary. If we find a way to make use of this information contained in the output, we might be able to circumvent the problem of having to postulate an input generator.

The information conveyed by Legendre et al.’s inputs resembles the syntactic structure we are assuming. The structures in (60) correspond to those in (58). The three verbal extended projections contained in these structures are labeled EP\(_1\), EP\(_2\), and EP\(_3\). Crucially, the position of the [+wh] feature differs in the two representations: in (60a), it is located in EP\(_2\) (on I), while in (60b), it is situated in EP\(_3\) (again on I). The simplest way to prevent the structures in (60) from competing with each other is by prohibiting the comparison of representations in which formal features are associated with different extended projections.

(60)  
\[ \text{EP}_1 \ldots \text{V} \ldots [\text{EP}_2 \ldots [+\text{wh}] \ldots \text{V} \ldots [\text{EP}_3 \ldots \text{V} \ldots]] \]

This follows from the principle in (61). In order to be able to verify if features have an identical distribution in two representations R\(_1\) and R\(_2\) (condition (61c)), each extended projection in R\(_1\) must have exactly one correspondent in R\(_2\), and vice versa (condition (61a)), and the structural relation between the two extended projections in R\(_1\) must be identical to the structural relation between their correspondents in R\(_2\) (condition (61b)). Let us assume that if extended projections have moved, only their base-position is relevant.
The two representations \( R_j \) and \( R_2 \) are members of one candidate set iff (a)-(c) holds.

a. Each extended projection in \( R_j \) has exactly one correspondent in \( R_2 \) and vice versa.

b. The structural relation between (the base-position of) two extended projections in \( R_j \) is identical to the structural relation between (the base-position of) their correspondents in \( R_2 \).

c. Each extended projection in \( R_j \) is composed of the same set of formal features as its correspondent in \( R_2 \).

Notice that according to condition (61), \( R_j \) and \( R_2 \) can be based on distinct numerations as long as these contain the same formal features.

Condition (61) is an illustration of the fact that next to the advantages of D&E over mainstream minimalist syntax, it is also fruitful from an OT perspective, since it enables us to solve the problem of candidate evaluation in a natural manner. Features are one of the driving forces behind movement. At the same time, they determine which structures may enter in competition with each other. Hence, we do not have to take recourse to structured inputs or semantic conditions on candidate sets, unlike alternative OT models of syntax.

Whether (61) is appropriate or not is largely an empirical matter. For instance, it follows from (61) that we are free to introduce lexical items that do not contain formal features in our candidate structures. In other words, members of a candidate set need not be composed of the same set of words. If formal features are those properties of lexical items that are involved in feature checking, elements which do not play a role in movement transformations, such as the complementizer \textit{that} in (62), are freely inserted.

(62) a. I think that the President has covered up the truth.

b. I think the President has covered up the truth.

In chapters 3 and 4, we will argue that CP is only projected in examples like (62) if the complementizer is pronounced. The relevant structures are given in (63). According to (61), these two structures are contained in one candidate set. The structure in (63a) consists of five extended projections (labeled \( EP_1, \ldots, EP_5 \)) which each have a correspondent in (63b). The structural relations between these five extended projections as well as their featural content are identical in both structures. There is only one difference between the two structures: \( EP_3 \) in (63a) contains a CP (with a pronounced head), whereas its correspondent in (63b) does not. However, the CP in (63a) is not involved in feature checking.\(^{21}\) Hence, its presence or absence in \( EP_3 \) is innocent in the light of (61).

\(^{21}\) Note that non-insertion and deletion are two distinct phenomena. In chapter 3, we will see that although elements that move or attract must be contained in each member of the candidate set, they can, in principle, be deleted (see also section 3.1 above, where relative pronouns are deleted after movement).
(63) a. \([\text{EP}_1 [\text{EP}_2 \text{I think that } \text{IP} [\text{EP}_4 \text{ the President}] \text{ has covered up } \text{EP}_5 \text{ the truth}]])\]
b. \([\text{EP}_1 [\text{EP}_2 \text{I think } \text{IP} [\text{EP}_4 \text{ the President}] \text{ has covered up } \text{EP}_5 \text{ the truth}]])\)

Also the two examples in (64) are in one candidate set, in accordance with the analysis put forth in chapter 1.

(64) a. Who read these books?
b. *Who did read these books?

The structures are given in (65). Each extended projection in (65a) has its correspondent in (65b). Relations between extended projections are identical in both structures. Furthermore, formal features are identically distributed over the extended projections. The fact that \(\text{EP}_1\) contains more words in (65b) than in (65a) is again irrelevant.

(65) a. \([\text{EP}_1 [\text{EP}_2 \text{who} \text{ read } \text{EP}_3 \text{ these books}]])\]

b. \([\text{EP}_1 [\text{EP}_2 \text{who} \text{ did read } \text{EP}_3 \text{ these books}]])\]

In sum, the examples in (62) and (64) provide evidence for the assumption that a candidate set need not be based on a single numeration. Furthermore, we will argue in chapter 4 that similar to \(\text{that}\), complementizers with clear semantic content (e.g. \(\text{if}\) in embedded yes-no questions) can freely be omitted. This, as we will see, is incompatible with a semantic identity requirement.

4.3. On the division of labor between Eval and Gen/C_HL

Broekhuis & Dekkers (to appear) argue that constraints should only be considered a member of the constraint set \(\text{Con}\) if they are violable. If a particular constraint is universally undominated, it should be reformulated as a principle that is operative in Gen/C_HL, the locus of universal linguistic statements. Pesetsky's REC (see section 3.1 above) might be such a constraint: in Pesetsky (1997, 1998), as well as in Broekhuis & Dekkers (to appear) and in the chapters 3-5 of this book, it always appears as the highest-ranked constraint. Broekhuis & Dekkers argue that it should be replaced by the condition on the operation Delete given in (45). Such a condition can be formulated along the lines of (66).

(66) Do not remove the phonological matrix of syntactic units with semantic content.

Another inviolable principle of language is that \(\text{wh}\)-movement does not apply from subjects or adjuncts, as illustrated in (67a) and (67b) respectively (see Huang, 1982). Movement from complements, on the other hand, is not prohibited, as shown in (67c). These and similar facts call for a theory of possible extraction domains (Bounding Theory). Although Bounding Theory has played a central role in
generative syntax since Ross (1967), Chomsky (1995: ch. 3-4) is silent about it (see however Zwart, 1997).

(67)  a. *Who are pictures of on sale?
b. *Who did Mary cry after John hit?
c. What do you think that John will do?

If the pattern in (67) is indeed universal, it seems likely that the application of the operation Move is restricted in Gen/C_{hl} by a subjacency condition such as the one in (68), which states that all maximal projections intervening between the source and the target of movement must belong to the same extended projection.

(68) If α moves to the minimal domain of X, every maximal projection dominated by XP and dominating α is a sister of an extended head of XP.

The structures in (69) correspond to the examples in (67a-b). In these structures, the wh-phrase moves to the minimal domain of C. ZP is dominated by CP, and it dominates $t_i$. Hence, ZP qualifies as an intervening maximal projection if it is not the sister of an extended head of CP. ZP is a subject in (69a) and an adjunct in (69b). Since neither subjects nor adjuncts are sisters of heads, (68) excludes both structures.

(69) a. [CP who$_j$ [C are] [IP [ZP pictures of $t_i$] on sale]]
b. [CP who$_j$ [C did] [IP Mary cry [ZP after John hit $t_i$]]]

In (70), on the other hand, ZP is the sister of the verb think, an extended head of CP. Therefore, ZP does not count as an intervening projection. However, if who moves to the matrix clause in one step, the lower IP and VP still count as intervening nodes because their sisters C and I are not extended heads of the matrix CP. Since C and I are extended heads of the lower CP, this can be remedied by moving who through the lower SpecCP. Thus, (68) implies successive cyclicity. Hence, if (68) is part of Gen/C_{hl}, all movement is successive cyclic.

(70) [CP what$_j$ [C do] [IP you think [ZP=CP $t_i$ that [IP John will [VP do $t_i$]]]]]

In sum, we should examine for each condition on movement or pronunciation whether or not it is violable. Until there is evidence for the violability of a particular condition, it should be formulated as a principle in Gen/C_{hl}.

5. Conclusion

In this chapter, we have argued in favor of D&E, an OT syntax in which the generator resembles Chomsky’s (1995) computational system C_{hl}. A generator in OT syntax should be simple and universal: two requirements met by C_{hl}. Furthermore, it should overgenerate in such a way that it produces candidate sets that
are rich enough to be successfully evaluated with respect to constraint rankings. \( \mathcal{C}_{HL} \) also qualifies in this respect, provided that the operation Move applies freely (contra Chomsky), and the operation Delete is added to the system. Although a generative grammar is by definition explicit about the way in which it produces outputs, little attention has been paid to the contents of the generative module in OT. This void is filled if the generator is identified with a Chomsky-style computational system. This step has several positive side-effects. For instance, it allows us to relate the existence of universally ungrammatical structures to specific properties of Gen/C\( \mathcal{L} \). Also, a feature-based generator obliterates the need for a structured input or a semantic identity requirement on candidate sets, both of which are problematic.

Compared to C\( \mathcal{L} \) as developed in Chomsky (1995), D&E allows us to circumvent any (indirect) reference to Spell-Out, and to unify Fewest Steps and Procrastinate. Furthermore, various syntactic phenomena which are difficult or impossible to capture in terms of feature strength, such as the influence of information structure on word order and complementizer pronunciation, receive a natural explanation in OT syntax such as D&E. By reformulating feature strength in terms of constraint interaction, a general constraint-based theory of syntactic parametrization with a wider empirical range and greater typological power than standard minimalist syntax can be constructed.

In chapters 3-5, aspects of the syntax of subjects and complementizers will be analyzed within D&E. The focus will lie on constraint interaction, since in principle all parametrizing power of D&E (and therefore a substantial part of its explanatory adequacy) is located in the evaluation procedure. Nevertheless, the specific properties we have attributed to the Gen/C\( \mathcal{L} \) play an implicit role throughout. They receive explicit attention whenever necessary.