The Rise of Competition in Syntax: A Synopsis

Gereon Müller & Wolfgang Sternefeld

1 Local vs. Competition-Based Approaches

Syntactic theories differ with respect to how they determine the wellformedness or illformedness of a given sentence S_i in a given language. One possibility is that the decision of whether S_i is grammatical or not can be made by exclusively considering properties of S_i; properties of other sentences S_j, S_k, ... are irrelevant. Another possibility is that properties of other sentences S_j, S_k, ... do play a role in deciding whether S_i is grammatical or not in addition to S_i's own properties. The first possibility, which we may call a *local approach*, can arguably be viewed as the standard one; this strategy is pursued in, e.g., most versions of government and binding theory (principles and parameters theory), head-driven phrase structure grammar, lexical-functional grammar (until recently), and in certain versions of the minimalist program. The second possibility presupposes a competition of sentences; hence, we will refer to it as a *competition-based approach*. This strategy is the one that this book is about; it is chosen in certain versions of the minimalist program (in particular, in earlier manifestations), in theories that incorporate the Blocking Principle (the Elsewhere Condition), and, last but not least, in optimality-theoretic syntax. In what follows, we will illustrate fundamental differences between and points of convergence among local and competition-based approaches by considering government and binding theory (section 2) and its development into the minimalist program (section 3), blocking syntax (section 4), and optimality-theoretic syntax (section 5).

2 Government and Binding Theory

Chomsky's (1981, 1986a,b) theory of government and binding is a typical instance of a local approach. In this theory, syntactic objects are viewed as

⟨D-structure, S-structure, LF⟩ triples which are created by phrase-structure rules (X-bar theory) and the transformational rule Move α. Syntactic constraints can take various forms. First, they can be representational filters in the sense that they apply at one or more of the three levels D-structure, S-structure and/or LF (cf. the principles of θ-theory, the binding theory principles, or the Empty Category Principle (ECP)). Alternatively, constraints can be derivational, which implies that they do not apply at any specific level, but restrict the movement operation itself (cf. the Subjacency Condition). Finally, government and binding theory envisages the possibility that syntactic constraints can be global (in Lakoff's 1971 terminology). Global constraints are neither representational nor derivational; rather, they relate non-adjacent representations in a complex derivation. In the case at hand, the Projection Principle (see Chomsky 1981:38) is a global constraint that relates the three levels of representation by demanding that the subcategorization properties of lexical items be respected from one level to the next. Still, all these types of constraints, including the global one, can be checked by exclusively looking at a given syntactic object S_i.[1] Properties of other syntactic objects are irrelevant. Thus, all these types of constraints (representational, derivational, and global) can be viewed as local in this sense.

2.1 Representational Constraints

To see this, consider first the effects of representational constraints like the binding principles A and B, which are given here in a simplified form.

(1) a.   *Principle A*:
          An anaphor is bound in its binding domain.
    b.   *Principle B*:
          A pronominal is free in its binding domain.

By assuming that these principles apply at S-structure, we can derive that (2-a,c) (more precisely, the ⟨D-structure, S-structure, LF⟩ triples of which (2-a,c) are simplified S-structure representations) are grammatical, whereas (2-b) and (2-d) are ungrammatical, due to violations of Principle A and Principle B, respectively. Crucially, however, the determination of grammaticality of the four sentences proceeds on a local basis. Thus, the illformedness of (2-d) is not related to the wellformedness of (2-a), and the illformedness of

(2-b) is not related to the wellformedness of (2-c), even though the two strate-gies – pronominalization and reflexivization – seem to be in complementary distribution for the most part in English.

(2) a.   John$_1$ likes himself$_1$
  b.   *John$_1$ thinks that Mary likes himself$_1$
  c.   John$_1$ thinks that Mary likes him$_1$
  d.   *John$_1$ likes him$_1$

Another representational constraint in government and binding theory is the ECP, which is assumed to apply to LF representations only (see Chomsky 1986a, Lasnik & Saito 1992).

(3)   *ECP* ("Empty Category Principle"):
      Every trace must be marked $[+\gamma]$.

A trace is marked $[+\gamma]$ if it is properly (antecedent or lexically) governed, and $[-\gamma]$ if it is not properly governed. Once assigned, a γ-feature is irreversible. By assumption, the complementizer *that* blocks antecedent-government of a subject trace. Consequently, t_1 is marked $[-\gamma]$ in (4-a), which exhibits a familiar complementizer-trace effect, and $[+\gamma]$ in (4-b). Even though these features are assigned at S-structure, the actual ECP violation in (4-a) takes place at LF. There is a reason why the ECP cannot be assumed to hold at S-structure already: If it did, (4-c) would also be ill formed, due to $[-\gamma]$-marking of the intermediate subject trace t'_1. Being an argument trace in an A-bar position, this intermediate trace is freely deletable on the way to LF (an option that does not exist for t_1 in (4-a,b,c)). With it goes the $[-\gamma]$-marking, and (4-c) is correctly predicted to be well formed despite a lack of proper government at S-structure – the ECP is respected at LF.

(4) a.   *Who$_1$ do you think [$_{CP}$ t$'_1$([$+\gamma$]) that [$_{IP}$ t$_1$([$-\gamma$]) will leave]] ?
  b.   Who$_1$ do you think [$_{CP}$ t$'_1$([$+\gamma$]) — [$_{IP}$ t$_1$([$+\gamma$]) will leave]] ?
  c.   Who$_1$ do you think [$_{CP}$ t$''_1$([$+\gamma$]) that Mary said [$_{CP}$ t$'_1$([$-\gamma$]) — [$_{IP}$ t$_1$([$+\gamma$]) will leave]]] ?

It is again worth noting that the account of the illformedness of (4-a) in no way relies on the fact that (4-b) is well formed – the grammaticality status of the two sentences is determined on a local basis.

2.2 Derivational Constraints

Basically the same situation arises with constraints like the Subjacency Condition in (5).

(5)   *Subjacency Condition*:
        Movement must not cross two bounding nodes.

The standard assumption in government and binding theory is that Subjacency is a derivational constraint on (overt) movement, rather than a representational constraint applying at S-structure. Theory-internal motivation for this assumption comes from the consideration of the data in (6) (cf. Sternefeld 1991).

(6) a.   Who$_1$ do [$_{IP}$ you think [$_{CP}$ t$'_1$ that [$_{IP}$ Mary loves t$_1$]]] ?
    b.   Who$_1$ do [$_{IP}$ you believe [$_{IP}$ John to be in love with t$_1$]] ?

IP is a bounding node for movement. The two *wh*-movement steps in (6-a) cross one IP each, and given that no other bounding node intervenes, the wellformedness of this sentence is compatible with both a derivational and a representational interpretation of the Subjacency Condition. The case is different in the exceptional Case marking context in (6-b). As it stands, t$_1$ is separated from its antecedent *who*$_1$ by two IP bounding nodes at S-structure, since there is no intermediate trace t$'_1$. Thus, under a representational interpretation of the Subjacency Condition, we would in fact wrongly expect (6-b) to be ungrammatical. No such problem arises under a derivational interpretation, though. On this latter view, (6-b) is derived from a D-structure representation with a CP shell present in the embedded clause. *Wh*-movement can precede CP deletion (as required for exceptional Case marking of the embedded subject NP), and the Subjacency Condition is correctly predicted to be satisfied by the movement operation itself. This is shown in (7), which outlines the relevant part of the derivation of (6-b).

(7) a.   — do [$_{IP}$ you believe [$_{CP}$ — C [$_{IP}$ John to be in love with who$_1$]]]
    b.   — do [$_{IP}$ you believe [$_{CP}$ who$_1$ C [$_{IP}$ John to be in love with t$_1$]]]
    c.   who$_1$ do [$_{IP}$ you believe [$_{CP}$ t$'_1$ C [$_{IP}$ John to be in love with t$_1$]]]
    d.   who$_1$ do [$_{IP}$ you believe [$_{IP}$ John to be in love with t$_1$]]

Among other phenomena, the Subjacency Condition derives Complex Noun Phrase Condition (CNPC) islands. The contrast between (8-a) and (8-b) follows, given that *wh*-movement crosses two bounding nodes in (8-b), but not

in (8-a). Again, the fact that the movement strategy seems to be in complementary distribution with a resumptive pronoun strategy is not theoretically reflected by relating the illformedness of resumptive pronouns as in (8-c) to the wellformedness of movement in (8-a), and the (relative) wellformedness of resumptive pronouns in (8-d) to the illformedness of movement in (8-b):[2] Independently of how exactly an account of the contrast in (8-c,d) looks like in government and binding theory, it seems that it must rely on a local constraint that is violated if a resumptive pronoun has an antecedent that is too close.[3]

(8) a.   the man [$_{CP}$ who(m)$_1$ I saw t$_1$]
  b.   *the man [$_{CP}$ who(m)$_1$ [$_{IP}$ I don't believe [$_{NP}$ the claim [$_{CP}$ t$'_1$ that anyone saw t$_1$]]]]
  c.   *the man [$_{CP}$ who(m)$_1$ I saw him$_1$]
  d.   ?the man [$_{CP}$ who(m)$_1$ [$_{IP}$ I don't believe [$_{NP}$ the claim [$_{CP}$ that anyone saw him$_1$]]]]

2.3 Global Constraints

Finally, the Projection Principle can always be checked by looking at the properties of a given sentence, and does not necessitate a consideration of other sentences. The Projection Principle is given in a simplified form in (9).

(9)   *Projection Principle*:
  a.   If α selects β in γ as a lexical property, then α selects β in γ at some level L$_i$.
  b.   If α selects β in γ at level L$_i$, then α selects β in γ at level L$_j$.

To find out whether a given sentence S$_i$ respects the Projection Principle, we have to take into account the sentence's representations at D-structure, S-structure, and LF. This way, the presence of traces of overt movement in argument positions can be enforced at S-structure and at LF, among other things. Clearly, such a global constraint is quite complex; however, it does not yet rely on the notion of competition – properties of other sentences (i.e., other ⟨D-structure, S-structure, LF⟩ triples) are irrelevant for determining whether a given sentence violates or respects the Projection Principle.

2.4   An Exception: The Avoid Pronoun Principle

To sum up so far, government and binding theory has different types of constraints, with varying complexity, but all of them are local in the above sense; i.e., they do not involve competition. Interestingly, though, there is a notable exception: the Avoid Pronoun Principle of Chomsky (1981).

   The empty pronominal PRO and lexical pronouns come close to being in complementary distribution: As stated in the PRO-theorem, PRO is confined to positions that are ungoverned (e.g., the subject position of control infinitives). In contrast, overt pronouns typically show up in positions that are governed. This is so because overt pronouns must be assigned Case, and Case is normally assigned under government only. However, there is one position in which government and binding theory permits both an empty pronominal PRO (because the position is ungoverned) and an overt pronoun (because Case can be assigned without government by a special Case assignment rule). This is the subject position of English gerunds. Consider first the possibility of PRO in this position:

(10) a.   John$_1$ would much prefer [PRO$_1$ going to the movie]
     b.   *John$_1$ would much prefer [PRO$_{2/arb}$ going to the movie]

As shown by the contrast between (10-a) and (10-b), PRO must be co-indexed with the matrix subject here; it cannot bear a different index or be interpreted arbitrarily. Obligatory control follows from the control rule in (11) (see Manzini 1983).[4]

(11)   *Control Rule*:
         If PRO is minimally dominated by a declarative clausal object α, then it must be controlled by an antecedent within the minimal CP dominating α.

Next consider overt pronouns in English gerunds:

(12) a.   *John$_1$ would much prefer [his$_1$ going to the movie]
     b.   John$_1$ would much prefer [his$_2$ going to the movie]
     c.   John$_1$ would much prefer [his$_1$ book]

(12-b) shows that an overt pronoun is possible in the subject position of English gerunds. However, an overt pronoun cannot be co-indexed with the matrix subject, in striking contrast to PRO; cf. (12-a). The illformedness of

(12-a) does not follow from any of the relevant local constraints in government and binding theory. In particular, (12-c) strongly suggests that principle B of the binding theory is not violated in (12-a) (the pronoun *his* occupies SpecNP in both cases, according to Chomsky's 1981 assumptions), and the wellformedness of (12-b) proves that Case can be assigned to *his* in (12-a). In view of this situation, Chomsky (1981:65) suggests that the illformedness of (12-a) is not to be traced back to a violation of some local constraint. Rather, it should be related to the wellformedness of (10-a): The two sentences compete, and (12-a) is ungrammatical because (10-a) is grammatical. This idea is implemented by adopting the Avoid Pronoun Principle in (13).

(13)   *Avoid Pronoun Principle*:
         Lexical pronouns are blocked by empty pronouns if possible.

This implies that the grammaticality of a sentence S_i with an overt pronoun cannot be checked on a local basis anymore: To find out whether such a sentence S_i is grammatical or not, a minimally different sentence S_j has to be considered in which the overt pronoun is replaced by PRO. If S_j is grammatical, S_i is ungrammatical by the Avoid Pronoun Principle; this effect occurs in the case of (10-a) vs. (12-a). If, on the other hand, S_j is ungrammatical (e.g., because PRO violates the PRO-theorem, or because PRO has an index that leads to a violation of the Control Rule), S_i can be grammatical (provided that no local constraints of grammar are violated); this situation arises in the case of (10-b) (which violates the local Control Rule) vs. (12-b).

The question arises of what type of constraint the Avoid Pronoun Principle is. Clearly, it does not belong to any of the three types of local constraints discussed above: To find out whether a given sentence S_i respects or violates this constraint, it does not suffice to consider properties of S_i alone. Rather, the properties of another sentence S_j must be considered. For the present purposes, let us refer to constraints like the Avoid Pronoun Principle that rely on a competition of sentences as *translocal* constraints.[5]

It seems that the Avoid Pronoun Principle has been widely accepted in government and binding theory. However, conceptually it has arguably always been an alien element in that syntactic approach. In line with this, it has never been fully clear whether it should best be viewed as a purely syntactic constraint, or indeed as a pragmatic constraint derivable from, e.g, Gricean maxims.[6] Such a question does not arise with the use of translocal constraints in the minimalist program.

3   The Minimalist Program

Translocal constraints are employed in various versions of the minimalist program that are developed in Chomsky (1991, 1993, 1995, 1998). In general, translocal constraints can be viewed as selection instructions: Out of a given set of syntactic objects, one (or more) is chosen according to a metric that is specified by the constraint. The set of syntactic objects that participate in the competition is called the reference set. The minimalist program dispenses with D-structure and S-structure, retaining only the interface levels LF and PF as levels of representation. Consequently, representational constraints that apply at D-structure or S-structure are dispensed with (as well as global constraints). The remaining local constraints are either derivational, or they apply at LF or PF ("bare output conditions"). With respect to the issue of competition, this derivational orientation of the minimalist program has an immediate consequence. The competing syntactic objects are derivations (not representations or n-tuples of levels of representations). In line with this, translocal constraints in the minimalist program are often called *transderivational* constraints. A common property of most (if not all) translocal constraints in the minimalist program is that they can be conceived of as *economy* constraints in some sense; i.e., a translocal constraint choses the most economical derivation in the reference set, according to some metric of economy. This certainly holds for the first translocal constraint suggested in the minimalist program, the Fewest Steps condition.

3.1   Fewest Steps

3.1.1   *V-to-I Movement in Chomsky (1991)*

Chomsky (1991) is concerned with deriving the difference between French and English with respect to V-to-I movement (cf. Pollock 1989): French has overt V-to-I movement of finite verbs; English does not have such movement (except for auxiliaries). This is shown in (14).

(14) a.   Jean embrasse$_1$ souvent [$_{VP}$ t$_1$ Marie]
      b.   *Jean souvent [$_{VP}$ embrasse$_1$ Marie]
      c.   *John kisses$_1$ often [$_{VP}$ t$_1$ Mary]
      d.   John often [$_{VP}$ kisses$_1$ Mary]

As a first step towards accounting for these data, Chomsky assumes that French has "strong" I nodes, whereas English has "weak" I nodes. This distinction becomes important for the following local (derivational) constraint:

(15)   *Strength of I*:
        Strong I tolerates adjunction of all Vs; weak I tolerates adjunction only of "light" Vs (auxiliaries).

This excludes (14-c) in English: Overt V-to-I movement violates Strength of I. In contrast, overt V-to-I movement in (14-a) in French does not violate this constraint. Still, something extra needs to be said about (14-b) in French, which vacuously fulfills Strength of I, just like (14-d) in English does. Thus, the question is: Why is overt V-to-I raising obligatory if it is possible? Chomsky's (1991) background assumption is that inflection is base-generated in I. If V does not raise to I, I must lower to V in overt syntax, so as to fulfill another local constraint, the *Stray Affix Filter*, which prohibits inflectional affixes that are not attached to a verbal host. The crucial idea now is that overt I lowering creates an unbound trace that must be undone by LF via covert raising of the whole V-I complex to the position of the trace of I. The derivations underlying (14-a) and (14-b) in French are given in (16) and (17), respectively.

(16) a.   Jean I_2 souvent [$_{VP}$ embrasse$_1$ Marie]   (overt raising)
     b.   Jean [$_{I_2}$ embrasse$_1$-I_2] souvent [$_{VP}$ t$_1$ Marie]

(17) a.   Jean I_2 souvent [$_{VP}$ embrasse$_1$ Marie]   (overt lowering)
     b.   Jean t$_2$ souvent [$_{VP}$ [$_{V_1}$ embrasse$_1$-I_2] Marie]   (covert raising)
     c.   Jean [$_{V_1}$ embrasse$_1$-I_2] souvent [$_{VP}$ t$_1$ Marie]

The second derivation has more movement steps than the first one, and it is therefore filtered out as uneconomical by the translocal economy constraint *Fewest Steps*, which can be formulated as follows:

(18)   *Fewest Steps*:
        If two derivations D_1 and D_2 are in the same reference set and D_1 involves fewer operations than D_2, then D_1 is to be preferred over D_2.

A definition of reference set that works for the approach in Chomsky (1991) is (19); here, the numeration is the set of all lexical items (including functional heads) that are used in a derivation.[7]

(19)   *Reference Set*:
        Two derivations D_1 and D_2 are in the same reference set iff they (i) have the same numeration and (ii) respect all local constraints.

The qualification in (i) ensures that, e.g., (20-b) cannot accidentally block (20-a) even though it involves fewer syntactic operations. Furthermore, the statement in (ii) produces the welcome consequence that (20-c) cannot accidentally block (20-a) even though it involves fewer syntactic operations by leaving the *wh*-phrase and the auxiliary in situ – (20-c) violates a local constraint like the *Wh-Criterion*, which requires a *wh*-phrase to move to $SpecC_{[+wh]}$ overtly in English.

(20) a. $What_1$ $have_2$ you t_2 seen t_1 ?
     b.   You have seen a car
     c.   *You $have_2$ seen wht_1 ?

This particular application of a translocal constraint in the minimalist program is not generally accepted anymore. Still, it can serve as an illustration of certain recurring properties and problems of translocal constraints, and indeed of competition-based approaches in general.

   First, as Chomsky (1991:433) observes, this system "tends to eliminate the possibility of optionality in derivation. Choice points will be allowable only if the resulting derivations are all minimal in cost ... This may well be too strong a conclusion, raising a problem for the entire approach." As an example, consider optional topicalization in English. (21-a) and (21-b) are both well formed, even though (21-b) invariably involves one movement operation more than (21-a).

(21) a.   Mary gave a book to $John_1$
     b.   To $John_1$ Mary gave a book t_1

The solution suggested by Chomsky (1991:433) is that certain movement operations (like, we can assume, topicalization in English) might be "assigned to some other component of the language system, perhaps a 'stylistic' component of the mapping ... to PF." Movement operations of this type might then be exempt from the Fewest Steps constraint. Alternatively, we might revise the definition of reference set appropriately, such that the two derivations in (21) do not compete anymore. For instance, we might add the requirement that competing derivations must have identical LF representations, as in (22).

Assuming that sentences which differ with respect to whether topicalization has applied must have different LFs, this would yield the desired result.

(22)   *Reference Set* (revised):
        Two derivations D_1 and D_2 are in the same reference set iff they (i) have the same numeration and the same LF, and (ii) respect all local constraints.

Second, a potential conceptual problem arises. Translocal economy constraints increase complexity: To find out whether a given sentence is grammatical, it does not suffice to look at internal properties of the sentence (Does it violate a local constraint?); rather, the properties of other sentences have to be taken into account as well (Does the sentence have the most economical derivation in the reference set?). Chomsky (1991:448) remarks that this might suggest that "language design as such appears to be in many respects 'dysfunctional,' yielding properties that are not well adapted to the functions language is called upon to perform."[8]

Third, Chomsky (1991) discusses successive-cyclic movement. It is standardly assumed that long-distance *wh*-movement of adjunct *wh*-phrases must be successive-cyclic; otherwise, a locality constraint (like the ECP) will be fatally violated, as with, e.g., *wh*-island configurations; cf. (23).

(23)   *How$_1$ do you wonder [$_{CP}$ whether to fix the car t$_1$] ?

Chomsky (1991) observes that successive-cyclic movement creates a potential problem for the Fewest Steps condition: Successive-cyclic movement as in (24-a) should always be blocked by one-swoop movement as in (24-b).

(24) a.   How$_1$ do you think [$_{CP}$ t$_1''$ that John said [$_{CP}$ t$_1'$ that Bill fixed the car t$_1$]] ?
     b.   *How$_1$ do you think [$_{CP}$ — that John said [$_{CP}$ — that Bill fixed the car t$_1$]] ?

Note, though, that no particular problem arises under the notion of reference set in (19) or (22). According to these definitions, only those derivations can compete that respect all local constraints of grammar, i.e., that are otherwise well formed. By hypothesis, the derivation that generates the surface representation (24-b) violates a locality constraint; hence, it cannot compete with the derivation that generates (24-a), and (24-a) is chosen by Fewest Steps because there is no competing derivation that would be more economical.

But what if we were to dispense with clause (ii) in the definition of reference set, or that clause (ii) were weakened in such a way that some derivations violating local constraints could compete after all. (As we will see below, there is some evidence for this latter option.) Then, the derivations generating (24-a) and (24-b) might compete, and the problem of accounting for successive-cyclic movement under Fewest Steps would persist. How, then, can we permit successive-cyclic movement in (24)? Chomsky (1993) advances the following solution: "Operations" as they are relevant for Fewest Steps do not simply involve applications of Move α as such. Rather, a more complex process of chain formation that (a) moves some item to its target position and (b) automatically inserts intermediate traces in appropriate positions counts as a single operation for the purposes of Fewest Steps:[9]

(25)   *Form Chain*:
        Move α to its target position and freely insert intermediate traces in appropriate positions.

On this view, "successive-cyclic" movement is no more costly from the perspective of Fewest Steps than one-swoop movement. Furthermore, the initial evidence concerning the French case of overt I-lowering followed by covert V-raising remains unaffected: A succession of movement operations involving a single item can only be reanalyzed as one instance of Form Chain (one operation for the purposes of Fewest Steps) if there is no other operation that intervenes; but in the case at hand, the operation spell-out that creates the branching in the derivation to PF and LF must intervene between overt lowering and covert raising in the derivation in (17). Hence, (17) still involves two applications of Form Chain.

3.1.2 *Wh-Topicalization in Epstein (1992)*

Another application of the Fewest Steps condition is the account of the ban on *wh*-topicalization in Epstein (1992). As noted above, topicalization is in principle optional in English; cf. (21). For many speakers, topicalization is also optionally possible in contexts like (26), where the target position is in an embedded clause and the matrix clause involves short *wh*-movement. Given the qualification that competing derivations must have identical LFs, this poses no problem for Fewest Steps.

(26) a.   Who$_1$ t$_1$ said [$_{CP}$ that [$_{IP}$ Mary gave a book to John$_2$]] ?
     b.   Who$_1$ t$_1$ said [$_{CP}$ that to John$_2$ [$_{IP}$ Mary gave a book t$_2$]] ?

Interestingly, embedded topicalization becomes impossible when the item that is topicalized is a *wh*-phrase; cf. (27-b). Note that the embedded *wh*-phrase may stay in situ in overt syntax, giving rise to a multiple question interpretation; cf. (27-a).

(27) a.   Who$_1$ t$_1$ said [$_{CP}$ that [$_{IP}$ Mary gave a book to whom$_2$]] ?

    b.   *Who$_1$ t$_1$ said [$_{CP}$ that to whom$_2$ [$_{IP}$ Mary gave a book t$_2$]] ?

Epstein proposes deriving the ban on *wh*-topicalization in (27-b) from the Fewest Steps condition. The derivations D$_1$ (generating (27-a)) and D$_2$ (generating (27-b)) are in the same reference set. Assuming that all *wh*-phrases must be in the domain of a SpecC$_{[+wh]}$ at LF, they both end up with the LF representation in (28):

(28)   *Who$_1$ to whom$_2$ t$_1$ said [$_{CP}$ that [$_{IP}$ Mary gave a book t$_2$]] ?

D$_1$ reaches this LF by applying one (covert) instance of *wh*-movement to the embedded object *to whom$_2$*. Note that there is only one movement operation in this case, either because LF movement of arguments does not have to be successive cyclic, or because successive-cyclic covert movement can be analyzed as one instance of Form Chain. D$_2$, on the other hand, reaches the same LF by applying two instances of *wh*-movement to the embedded object *to whom$_2$* (viz., one overtly and one covertly – given the intervening spell-out operation, these two movement operations cannot be reanalyzed as one instance of Form Chain). Hence, D$_1$ blocks D$_2$ via Fewest Steps.

   As shown in Müller & Sternefeld (1996), the same kind of analysis may be given for the ban on *wh*-scrambling in German, which is illustrated in (29).

(29) a.   Warum$_1$ hat der   Fritz was$_1$ gelesen ?

       why   has ART Fritz what read

    b.   *Warum$_1$ hat was$_1$ der   Fritz t$_1$ gelesen ?

       why   has what ART Fritz   read

However, it is also shown in Müller & Sternefeld (1996) that the Fewest Steps approach to the ban on optional movement of *wh*-phrases which is later undone by further, covert operations is not entirely unproblematic, and may necessitate additional assumptions. For one thing, German exhibits the same ban on *wh*-topicalization as English:

(30) a.   Wer$_1$ sagte t$_1$ [$_{CP}$ daß Maria wem$_2$ ein Buch gegeben hat$_3$] ?

       who said   that Maria whom a   book given   has

b. *Wer$_1$ sagte t$_1$ [$_{CP}$ wem$_2$ hat$_3$ Maria t$_2$ ein Buch gegeben t$_3$] ?
  who said   whom has Maria a book given

This strongly suggests an identical account in terms of Fewest Steps. However, since German topicalization always requires V/2 movement, and since V/2 movement is incompatible with the presence of a complementizer in German, the derivations generating (30-b) and (30-a) do not share an identical numeration, and we would wrongly expect no competition to arise.[10] Thus, to accomodate this evidence, it seems as though the definition of reference set must be revised, as in (31) – given complelementizer deletion at LF, (30-a,b) can be assumed to be identical at this level.

(31) *Reference Set* (second revision):
    Two derivations D$_1$ and D$_2$ are in the same reference set iff they (i) have the same LF, and (ii) respect all local constraints.

Moreover, it turns out that there are several well-formed constructions attested in the world's languages in which *wh*-phrases can in fact undergo optional overt fronting to a non-target position. In Müller & Sternefeld (1996), we discuss evidence from partial *wh*-movement, *wh*-imperatives, and *wh*-reconstruction. For the present purposes, the example of optional partial *wh*-movement to a SpecC$_{[-wh]}$ position in Ancash Quechua may suffice (cf. Cole 1982). (32-a) shows that *wh*-phrases may be fronted to a SpecC$_{[+wh]}$ target position in overt syntax in Ancash Quechua; (32-d) shows that *wh*-phrases may also stay in situ in overt syntax, raising (by assumption) to the SpecC$_{[+wh]}$ position in covert syntax. Interestingly, (32-b) and (32-c) are also possible. Here, the *wh*-phrases raise to an intermediate SpecC$_{[-wh]}$ overtly. Given that this implies an additional *wh*-movement operation at LF, Epstein's (1992) Fewest Steps approach should rule out these cases.

(32) a. [$_{CP}$ Ima-ta-taq$_1$ (qam) kreinki [$_{CP}$ t$_1''$ Maria muna-nqa-n-ta [$_{CP}$ t$_1'$
      what$_{acc}$ you believe   Maria want-nom-3-acc
    José t$_1$ ranti-na-n-ta]]] ?
    José buy-nom-3-acc
  b. [$_{CP}$ — (Qam) kreinki [$_{CP}$ ima-ta-ta$_1$ Maria muna-nqa-n-ta [$_{CP}$ t$_1'$
    José t$_1$ ranti-na-n-ta]]] ?
  c. [$_{CP}$ — (Qam) kreinki [$_{CP}$ — Maria muna-nqa-n-ta [$_{CP}$ ima-ta-ta$_1$
    José t$_1$ ranti-na-n-ta]]] ?
  d. [$_{CP}$ — (Qam) kreinki [$_{CP}$ — Maria muna-nqa-n-ta [$_{CP}$ — José ima-ta-ta$_1$ ranti-na-n-ta]]] ?

The conclusion drawn in Müller & Sternefeld (1996) in view of well-formed constructions like this one is that reference sets should be significantly reduced in size by assuming that identity of surface structure (rather than identity of LF) matters in the definition of reference sets, as in (33).

(33) *Reference Set* (third revision):
      Two derivations D_1 and D_2 are in the same reference set iff they (i) have the same surface structure, and (ii) respect all local constraints.

This way, partial *wh*-movement is permitted, but it is clear that much (in fact, most) of the original evidence in favor of Fewest Steps is lost: Thus, on this view, neither French V-in situ, nor English (or German) *wh*-topicalization can be ruled out by Fewest Steps anymore. As noted in Sternefeld (1997), this situation might be viewed as indicative of a general problem with translocal constraints: A significant reduction of competition in reference sets may be empirically desirable so as to account for cases of optionality (as in partial *wh*-movement constructions); but as an unwanted side effect, it also threatens to undermine the notion of translocal economy itself: Many ill-formed derivations that could be ruled out by translocal constraints will now survive because the more economical derivation is not part of the same reference set anymore. Finding a suitable definition of reference set that is weak enough to permit optionality and strong enough to actually do some work is one of the fundamental concerns of all versions of the minimalist program that employ the notion of competition.

3.1.3 *Freezing in Collins (1994)*

Evidence for yet another definition of reference sets comes from Collins' (1994) account of freezing effects with A-movement in English. As shown in (34-a,b), subject NPs are islands for extraction in English, whereas object NPs permit extraction (with certain types of verbs). In the present context, the interesting case is that of subject NPs that originate in object position, as in the case of passivization. As can be seen in (34-c), such derived subject NPs are also islands.

(34) a. Who_1 did John take [$_{NP}$ a picture of t_1] ?
     b.   *Who_1 is [$_{NP}$ a picture of t_1] on sale ?
     c.   *Who_1 was [$_{NP_2}$ a picture of t_1] taken t_2 by John ?

In a derivational approach, it must be shown that any derivation of (34-c) leads to illformedness. In one derivation, D_1, NP raising to subject position applies before *wh*-extraction from NP takes place.

(35) a. $[_{CP}$ — was $[_{IP}$ — taken $[_{NP_2}$ a picture of who$_1$] by John]]
   b. $[_{CP}$ — was $[_{IP}$ $[_{NP_2}$ a picture of who$_1$] taken t$_2$ by John]]
   c. *$[_{CP}$ who$_1$ was $[_{IP}$ $[_{NP_2}$ a picture of t$_1$] taken t$_2$ by John]]

Here, extraction of *who*$_1$ from NP$_2$ (which is in subject position already, hence non-L-marked, and therefore a barrier) violates a local constraint like the CED.[11]

(36)   *CED* ("Condition on Extraction Domain"):
      Movement must not cross a barrier.

In another derivation, D_2, *wh*-movement precedes NP raising:

(37) a. $[_{CP}$ — was $[_{IP}$ — taken $[_{NP_2}$ a picture of who$_1$] by John]]
   b. $[_{CP}$ who$_1$ was $[_{IP}$ — taken $[_{NP_2}$ a picture of t$_1$] by John]]
   c. $[_{CP}$ who$_1$ was $[_{IP}$ $[_{NP_2}$ a picture of t$_1$] taken t$_2$ by John]]

This derivation violates another local constraint, the Strict Cycle Condition in (38). The reason is that NP raising targets the subject position. The subject position is included in the CP domain, which has already been affected by *wh*-movement to SpecC earlier in the derivation.

(38)   *SCC* ("Strict Cycle Condition"):
      No movement operation may target a landing site that is included in a domain that has already been affected by movement earlier in the derivation.

So far, translocal constraints are not needed in an account of the illformedness of (34-c). The Fewest Steps condition does become relevant, though, when we consider a third derivation, D_3. This derivation proceeds by what Collins (1994) calls *chain interleaving*. First, the *wh*-phrase *who*$_1$ is extracted from the object NP$_2$ while NP$_2$ is still in situ (i.e., transparent for extraction); *who*$_1$ adjoins to VP. Second, NP raising to the subject position takes place. Finally, *who*$_1$ moves from its intermediate position to SpecC; see (39).

(39) a. $[_{CP}$ — was $[_{IP}$ — taken $[_{VP}$ $[_{NP_2}$ a picture of who$_1$] by John]]]
   b. $[_{CP}$ — was $[_{IP}$ — taken $[_{VP}$ who$_1$ $[_{VP}$ $[_{NP_2}$ a picture of t$_1$] by John]]]]

    c. $[_{CP}$ — was $[_{IP}$ $[_{NP_2}$ a picture of t_1] taken $[_{VP}$ who$_1$ $[_{VP}$ t_2 by John]]]]

    d. $[_{CP}$ who$_1$ was $[_{IP}$ $[_{NP_2}$ a picture of t_1] taken $[_{VP}$ t'_1 $[_{VP}$ t_2 by John]]]]

D_1 violates the CED; D_2 violates the SCC. D_3 violates neither of these local constraints. However, D_3 is blocked by D_1 and D_2 via Fewest Steps: Other things being equal, D_3 needs three movement steps where D_1 and D_2 make do with two movement steps.[12]

  This approach has an important consequence for the definition of reference sets. The three derivations D_1, D_2, and D_3 yield the same surface string, which is ill formed. Thus, the more economical derivations that block D_3 via Fewest Steps are not well-formed derivations, as in the applications of Fewest Steps discussed above, but rather ill-formed derivations that violate local constraints, viz., the CED and the SCC. This reasoning implies that reference sets can in fact not be defined as assumed so far, by requiring that only those derivations can compete that satisfy all local constraints – in the case at hand, D_1 and D_2 violate local constraints. Still, we cannot simply drop this requirement in the definition of reference sets; otherwise, all instances of movement would invariably be blocked in favor of in-situ derivations by Fewest Steps, and syntactic derivations would be fairly trivial. It seems that what is needed in view of this conflicting evidence is a relativized notion of local constraint satisfaction.

  In this context, the idea of *convergence* of derivations introduced in Chomsky (1993) becomes relevant: Only those derivations that converge can compete with respect to translocal constraints. Essentially, whereas all violations of local constraints lead to ungrammaticality, only a subset of violations of local constraints also leads to non-convergence. Ungrammatical derivations that converge may then still be used to block other derivations as ungrammatical, as in the freezing construction discussed by Collins (1994). It is an empirical issue how convergence is to be defined. As a rule of thumb, and for the present purposes, we can say that a violation of those constraints that trigger movement (like the *Wh*-Criterion, the Extended Projection Principle (EPP), which triggers subject raising, or whatever constraint optionally triggers topicalization) leads to non-convergence, whereas a violation of constraints like the CED and the SCC permits convergence of a derivation.[13]

  Under these assumptions, the notion of reference set needed for the approach in Collins (1994) can be defined as in (40). Note that the analysis is compatible with assuming that either numerations, or surface structures, or LF representations, or any combination of these determines the competition;

hence, a commitment to one of these options is not necessary in the case at hand.

(40)   *Reference Set* (fourth revision):
         Two derivations D_1 and D_2 are in the same reference set iff they (i) have the same numeration/surface structure/LF, and (ii) converge.

To end the discussion of Fewest Steps, we would like to emphasize that there is no inherent reason why the notion of an "operation" that is mentioned in the Fewest Steps condition should be confined to *movement*. Indeed, Chomsky & Lasnik (1993) argue that the *deletion* of intermediate traces in the LF component (which is argued to be an option with arguments and impossible with adjuncts in Chomsky 1986a, Lasnik & Saito 1992, and related work) is also regulated by the Fewest Steps condition.

There have been many more applications of the Fewest Steps condition in the minimalist program (see, e.g., the Fewest Steps account of the ban on semantically vacuous quantifier raising in Fox 1995), but these may suffice for the time being.[14] Let us now consider the translocal economy constraint Shortest Paths.

3.2   Shortest Paths

The Shortest Paths condition can be defined as follows (cf. Chomsky 1993, 1995):

(41)   *Shortest Paths*:
         If two derivations D_1 and D_2 are in the same reference set and the movement paths of D_1 are shorter than the movement paths of D_2, then D_1 is to be preferred over D_2.

Various applications of this condition have been suggested in minimalist syntax. Perhaps the most striking one concerns the derivation of superiority effects.

3.2.1   *Superiority Effects in Chomsky (1993) and Kitahara (1993)*

Superiority effects in English are illustrated by the examples in (42).

(42) a.   I wonder [$_{CP}$ who$_1$ C [$_{IP}$ t$_1$ bought what$_2$]]
     b.   *I wonder [$_{CP}$ what$_2$ C [$_{IP}$ who$_1$ bought t$_2$]]

    c.   Whom$_1$ did John persuade t$_1$ [$_{CP}$ to visit whom$_2$] ?

    d. *Whom$_2$ did John persuade whom$_1$ [$_{CP}$ t$'_2$ to visit t$_2$] ?

The Superiority Condition proposed by Chomsky (1973) demands that in cases where there are two (or more) *wh*-phrases that could in principle be moved to a given SpecC$_{[+wh]}$ position, only the highest *wh*-phrase can undergo such *wh*-movement overtly, i.e., the one that asymmetrically c-commands the other(s). This condition is respected in (42-a,c). The examples in (42-b,d) are ungrammatical because the highest *wh*-phrase in the clause has failed to undergo overt movement to SpecC; rather, the lower *wh*-object has moved to SpecC.

As indicated in Chomsky (1993) and argued extensively in Kitahara (1993), superiority effects can systematically be accounted for by the translocal condition Shortest Paths. For instance, the movement path from t$_1$ to *whom*$_1$ in (42-c) is shorter than the movement path from t$_2$ to *whom*$_2$ in (42-d). Hence, given that the two derivations D$_1$ and D$_2$ generating (42-c) and (42-d), respectively, compete (which would follow from most definitions of reference sets envisaged above), D$_1$ blocks D$_2$ as ungrammatical by Shortest Paths.

Or does it? Recall that at least some of the evidence for Fewest Steps (the ban on V-in situ in French, the ban on *wh*-topicalization and *wh*-scrambling in English and German) has relied on the assumption that covert movement counts in the same way that overt movement does. But assuming that LF movement also counts for the Shortest Paths condition leads straightforwardly into a dilemma: In the case at hand, the derivation that has the shorter overt *wh*-movement path invariably has the longer covert *wh*-movement path, and it seems that by LF, both derivations have *wh*-movement paths of equal length. Hence, ceteris paribus, both should be well formed. In view of this, several steps can be taken. First, one can assume that there is in fact no covert *wh*-movement of any kind; this makes it possible to maintain the Shortest Paths account of superiority phenomena without qualification, but is incompatible with the Fewest Steps applications sketched above. Second, one might explicitly distinguish between Fewest Steps and Shortest Paths in this respect: Whereas Fewest Steps compares whole derivations, Shortest Paths compares only the overt parts of derivations. A third possibility is developed in Sternefeld (1997).

The underlying intuition of this account is that LF movement from a position α to another position β has a chance to be shorter than overt movement from α to β. The central observation is that the issue of whether a Shortest Paths account of superiority effects is incompatible with covert *wh*-movement

is highly dependent on how path length is defined. If (movement) path length is determined by considering the number of nodes crossed by a movement operation, there is indeed a problem. But suppose now that path length is determined by considering the number of complete *chains* that are crossed. Now LF movement of *whom*$_2$ in (42-c) will create a shorter path than overt movement of *whom*$_2$ in (42-d), even though *whom*$_2$ originates in the same position and targets the same landing site. The reason is that covert movement of *whom*$_2$ in (42-c) crosses t_1, which is part of a chain, but not a complete chain, whereas overt movement of *whom*$_2$ in (42-d) crosses *whom*$_1$, which is a complete chain at this point of the derivation. Given that *whom*$_1$ crosses the same number of complete chains in the course of overt movement in D_1 (generating (42-c)) and in the course of covert movement in D_2 (generating (42-d)), the small difference pertaining to movement of *whom*$_2$ becomes decisive, and D_1 successfully blocks D_2 as ungrammatical via Shortest Paths, even under the assumption that covert *wh*-movement exists and is relevant for the Shortest Paths condition.

Another interesting issue raised by the Shortest Paths account of superiority effects is posed by what one might call "LF-optionality." Sentences like (43-a) have two possible readings (see Baker 1970) that correspond to two different LF representations, given LF movement of *wh*-in situ elements.

(43)   Who$_1$ t_1 wonders [$_{CP}$ where$_2$ we bought what$_3$ t_2] ?

    a.   who$_1$ what$_3$ t_1 wonders [$_{CP}$ where$_2$ we bought t_3 t_2]
         Answer: John wonders where we bought the books, Mary wonders
             where we bought the records, etc.

    b.   who$_1$ t_1 wonders [$_{CP}$ where$_2$ what$_3$ we bought t_3 t_2]
         Answer: John wonders where we bought what, Mary wonders
             where we bought what, etc.

Given that all *wh*-in situ phrases must undergo movement to a SpecC$_{[+wh]}$ position at LF, D_2 (creating (43-b)) should block D_1 (creating (43-a)) because D_1's paths are longer. Again, there are several possible solutions.[15] As before, one might stipulate that covert *wh*-movement either does not exist, or does not count with respect to the Shortest Paths condition. Alternatively, this evidence could be viewed as a further argument that reference sets are defined in such a way that competing derivations must have identical LF representations.

3.2.2 *Yo-Yo Movement in Collins (1994)*

The term *yo-yo movement* characterizes a combination of lowering and raising operations affecting a single item in the course of a derivation, or even within the overt part of a derivation. Derivations employing yo-yo movement are identified as problematic in Chomsky (1986a) (the main observation being attributed to Andy Barss), but envisaged as legitimate possibilities in Lasnik & Saito (1992). Collins (1994) shows that the availability of yo-yo movement would make a wrong prediction for the West African language Ewe, and attempts to derive a ban on yo-yo movement from the Shortest Paths condition.

Ewe is among the languages that show reflexes of successive-cyclic *wh*-movement in the C domain. The reflex of successive cyclicity concerns the morphological form of the 3.Pers.Sing. subject pronoun in the canonical subject position. The regular form of the pronoun is *é*; cf. (44-a). The regular pronoun *é* can be replaced by *wo* in cases of long-distance extraction (focus movement, in the case at hand); cf. (44-b).

(44) a.   Kofi gblŏ [$_{CP}$ be   é/*wo ʃo Kŏsi]
          Kofi said   that he   hit Kŏsi

     b.   Kofi$_1$ ε   me gblŏ [$_{CP}$ (t$'_1$) be   é/wo ʃo t$_1$]
          Kofi Foc I   said   that he   hit

Collins assumes as the correct underlying generalization that *é* is replaced by *wo* if and only if the local SpecC position is filled. Accordingly, we can postulate that the apparent optionality of *wo* in (44-b) is due to an option for long-distance A-bar movement of arguments in Ewe to apply either successive-cyclically, via SpecC (in which case *wo* is obligatory), or in one swoop (in which case *wo* is impossible). It does not come as a surprise from a pretheoretical point of view that A-bar movement that originates in the matrix clause and targets a SpecC position there does not trigger the morphological change in the embedded subject position:

(45)   Kofi$_1$ ε   me gblŏ na t$_1$ [$_{CP}$ be   é/*wo ʃo Kŏsi]
          Kofi Foc I   said to   that he   hit Kŏsi

Still, to ensure that *wo* is impossible in (45), a derivation like (46) that employs yo-yo movement must be ruled out. In this derivation D$_1$, *Kofi$_1$* is first lowered to the embedded SpecC position, licensing *wo* in the subject position there, and then raised to the target SpecC position in the matrix clause.[16]

(46) a.   Foc [$_{IP}$ I [$_{VP}$ said [$_{PP}$ to Kofi$_1$] [$_{CP}$ that [$_{IP}$ he hit Kŏsi]]]]

    b.   Foc [$_{IP}$ I [$_{VP}$ said [$_{PP}$ to t_1] [$_{CP}$ Kofi$_1$ that [$_{IP}$ he hit Kŏsi]]]]

    c.   Kofi$_1$ Foc [$_{IP}$ I [$_{VP}$ said [$_{PP}$ to t_1] [$_{CP}$ t_1' that [$_{IP}$ he hit Kŏsi]]]]

There are various possibilities to exclude such a derivation (relying, e.g., on versions of the SCC or versions of the Proper Binding Condition). Still, Collins (1994) observes that D_1 in (46) is blocked by the derivation D_2 in (47) via Shortest Paths. D_2 proceeds without intermediate lowering.[17]

(47) a.   Foc [$_{IP}$ I [$_{VP}$ said [$_{PP}$ to Kofi$_1$] [$_{CP}$ that [$_{IP}$ he hit Kŏsi]]]]

    b.   Kofi$_1$ Foc [$_{IP}$ I [$_{VP}$ said [$_{PP}$ to t_1] [$_{CP}$ that [$_{IP}$ he hit Kŏsi]]]]

A final remark is due concerning the notion of reference set presupposed by this analysis. Since "he" is a *wo* in D_1 and an *é* in D_2, it is clear that this difference must not suffice to create different reference sets. This can be accomplished in a number of ways, e.g., by defining reference sets with respect to a level of representation at which the difference in pronoun shape is invisible (possibly LF), or by explicitly stipulating in the definition of reference set that minor differences like the one at hand do not suffice to create different competitions. It is this latter strategy that is also pursued by Nakamura (1998) in his approach to *wh*-movement in Tagalog.

3.2.3 *Tagalog Wh-Movement in Nakamura (1998)*

A generalization underlying *wh*-movement in the Austronesian language Tagalog is that only the highest A-position of a given clause (the subject position) is accessible for *wh*-movement; Nakamura (1998) assumes this to be SpecT (or SpecI). In constructions in which an agent NP occupies the highest A-position (the so-called *Agent Topic* (AT) constructions), this NP can be *wh*-moved; an NP bearing a different Theta-role that shows up in an object position cannot undergo *wh*-movement; cf. (48).[18]

(48) a.   [$_{CP}$ Sino$_1$ ang [$_{TP}$ t_1' b-um-ili [$_{VP}$ t_1 t_V ng damit$_2$]]] ?
             who   Ang   bought$_{AT}$ dress$_{inh}$
             'Who is the one that bought the dress?'

    b.   *[$_{CP}$ Ano$_2$ ang [$_{TP}$ *si Juan*$_1$ b-um-ili [$_{VP}$ t_1 t_V t_2]]] ?
             what Ang   Juan$_{abs}$ bought$_{AT}$
             'What is the thing that Juan bought?'

A different marking on the verb triggers the so-called *Theme Topic* (TT) construction. Here, the theme NP occupies the structural subject position SpecT; and indeed, only the theme NP can undergo *wh*-movement; cf.:

(49) a. *[$_{CP}$ Sino$_1$ ang [$_{TP}$ *damit*$_2$ b-in-ili   [$_{VP}$ t$_1$ t$_V$ t$_2$]]] ?
           who Ang   dress$_{abs}$ bought$_{TT}$
         'Who is the one that bought the dress?'
     b. [$_{CP}$ Ano$_2$ ang [$_{TP}$ t$_2'$ b-in-ili   [$_{VP}$ ni Juan t$_V$ t$_2$]]] ?
           what Ang   bought$_{TT}$ Juan$_{erg}$
         'What is the thing that Juan bought?'

Nakamura's (1998) basic idea is that the derivations generating (48-a) and (49-a) compete, as do the derivations generating (48-b) and (49-b). The derivations underlying (48-a) and (49-b) can then block their respective competitors as ungrammatical because of the Shortest Paths constraint. To see this, consider the case of *wh*-movement of the theme NP in (48-b) and (49-b). The movement path from the VP-internal object position to the SpecC target position in (48-b) is longer than the path from the subject position SpecT to SpecC in (49-b). Consequently, the Shortest Paths condition guarantees that the derivation generating (49-b) blocks the derivation generating (48-b) as ungrammatical. An analogous account is available for the agent *wh*-movement case in (48-a) vs. (49-a).

   As Nakamura observes, this analysis raises two further potential problems. First, we have to ensure that derivations can compete even though they do not have identical lexical material – the Agent Topic and the Theme Topic constructions clearly differ in lexical make-up. Nakamura accomplishes this by replacing the notion of "identical numeration" in the definition of reference set with the more liberal notion of "non-distinct numeration;" the latter is defined in such a way that two numerations that only differ with respect to functional features do not count as distinct. (Clearly, this raises some non-trivial questions for other languages in which competitions of the type that Nakamura postulates seem unwanted.)

   Second, the derivation that generates, e.g., (49-b) may minimize the *wh*-path in comparison with the derivation that generates (48-b), but it increases path lengths in the A-domain. It is not quite clear how problematic this is; in the case presently under consideration, the A-chain formed in (49-b) by theme raising is only minimally longer than the A-chain formed in (48-b) by agent raising, whereas the *wh*-chain formed in (49-b) is much shorter than the *wh*-chain formed in (48-b). There would be even less of a problem for the agent *wh*-extraction case in (48-a) and (49-a). In any event, Nakamura (1998) replaces the notion of "movement paths" in the definition of the Shortest Paths condition with the more specific notion of "comparable chain links." This yields the effect that, e.g., the derivation generating (49-b) blocks the derivation generating (48-b) just because the former derivation's *wh*-chain

links are shorter than the latter derivation's comparable *wh*-chain links, irrespective of the length of other chain links created by A-movement, V raising, etc.

3.2.4 *Freezing in Chomsky (1995)*

Recall the English freezing construction in (34-c), which is repeated here:

(50)   *Who$_1$ was [$_{NP_2}$ a picture of t$_1$] taken t$_2$ by John ?

Above, we have considered three derivations. If NP raising precedes *wh*-movement, the CED is violated. If *wh*-movement precedes NP raising, the SCC is violated; the CED and the SCC are both local constraints. Finally, if chain interleaving occurs, this derivation can be excluded by invoking the Fewest Steps condition (cf. Collins 1994). Chomsky (1995) suggests that translocal economy constraints might play an even bigger role in accounting for the illformedness of (50). The idea is that the second derivation can be excluded without recourse to something like the SCC; the Shortest Paths condition can do this just as well. Consider again the two derivations that are compatible with Fewest Steps:

(51) a.   (i)   [$_{CP}$ — was [$_{IP}$ — taken [$_{NP_2}$ a picture of who$_1$] by John]]
          (ii)   [$_{CP}$ — was [$_{IP}$ [$_{NP_2}$ a picture of who$_1$] taken t$_2$ by John]]
          (iii) *[$_{CP}$ who$_1$ was [$_{IP}$ [$_{NP_2}$ a picture of t$_1$] taken t$_2$ by John]]
      b.   (i)   [$_{CP}$ — was [$_{IP}$ — taken [$_{NP_2}$ a picture of who$_1$] by John]]
          (ii)   [$_{CP}$ who$_1$ was [$_{IP}$ — taken [$_{NP_2}$ a picture of t$_1$] by John]]
          (iii)   [$_{CP}$ who$_1$ was [$_{IP}$ [$_{NP_2}$ a picture of t$_1$] taken t$_2$ by John]]

Chomsky's (1995:328) suggestion reads as follows: "Passive [i.e., NP raising] is the same in both [derivations]; *wh*-movement is 'longer' in the illicit one in an obvious sense, object being more remote from SpecC than subject in terms of number of XPs crossed. The distinction might be captured by a proper theory of economy of derivation." In other words: D$_2$ in (51-b), which does not violate a local constraint (the SCC, by assumption, being irrelevant or dispensable), is blocked as ungrammatical via Shortest Paths by the more economical D$_1$ in (51-a), which does violate a local constraint (the CED) but converges.

3.3 Procrastinate

Chomsky (1993, 1995) assumes the following local condition as a trigger for overt movement.

(52)   *Feature Condition*:
   a.   Strong features must be checked in overt syntax.
   b.   Weak features must be checked by LF.

As in the approach in Chomsky (1991), French I features are classified as "strong," English I features as "weak." However, whereas the constraint Strength of I (cf. (15)) in the 1991 analysis states that strong I *tolerates* V raising, the Feature Condition *forces* V raising to a strong I. Another difference between the two analyses is that the 1993 model does not require I lowering in the syntax anymore if V does not overtly raise to I. Looking back at the French/English paradigm in (14), we can now see that the original problem has disappeared: (53-a) respects the Feature Condition, (53-b) does not, and there is no need to invoke a translocal economy constraint to exclude (53-b).

(53) a.   Jean embrasse$_1$ souvent [$_{VP}$ t$_1$ Marie]
   b.   *Jean souvent [$_{VP}$ embrasse$_1$ Marie]
   c.   *John kisses$_1$ often [$_{VP}$ t$_1$ Mary]
   d.   John often [$_{VP}$ kisses$_1$ Mary]

However, a new problem has appeared: Assuming LF V-to-I movement in (53-d), this derivation respects the Feature Condition; but it seems that (53-c), with overt V-to-I movement, does so as well. This derivation was ruled out by Strength of I in Chomsky's (1991) approach, but with this condition gone, something else must be said. More generally, a condition is needed that guarantees that overt movement is possible only if it is forced by the Feature Condition, i.e., in the presence of strong features. This is achieved by the Procrastinate condition, which is explicitly formulated as a translocal constraint in Marantz (1995:357).

(54)   *Procrastinate*:
        If two derivations D_1 and D_2 are in the same reference set, and D_1 differs from D_2 in that an item α is moved covertly in D_1 and overtly in D_2, then D_1 is to be preferred over D_2.

Procrastinate blocks (53-c) in favor of (53-d), which delays V-to-I movement to the LF component. However, (53-b) does not block (53-a), assuming that only those derivations can compete that converge, and that (53-b) does not converge because of its Feature Condition violation.

The current status of Procrastinate is somewhat unclear. There have been attempts to dispense with this translocal constraint, either by deriving its effects from local constraints, or by reducing it to the Fewest Steps condition (cf. Kitahara 1997).[19]

3.4 Merge before Move

Chomsky (1995, 1998) assumes that syntactic structures are created by alternating operations of structure-building (Merge) and movement (Move). At any given stage of the derivation, the situation can arise that it must be decided whether the next step is a Merge or a Move operation. The following translocal condition settles the issue by preferring Merge to Move if both are possible as such; the specific formulation is based on Frampton & Gutman (1999).

(55) *Merge before Move*:
Suppose that two derivations D_1 and D_2 are in the same reference set and respect all local constraints, and $D_1 = \langle \Sigma_0, ..., \Sigma_n, \Sigma_{n+1}, ... \Sigma_k \rangle$ and $D_2 = \langle \Sigma_0, ..., \Sigma_n, \Sigma'_{n+1}, ... \Sigma'_k \rangle$. Then D_1 is to be preferred over D_2 if $\Sigma_n \rightarrow \Sigma_{n+1}$ is an instance of Merge and $\Sigma_n \rightarrow \Sigma'_{n+1}$ is an instance of Move.

Evidence for this condition comes from expletive constructions in English. Consider the data in (56-a,b).

(56) a. There$_1$ seems [$_{IP}$ t$_1$ to be [$_{PP}$ someone$_2$ in the room]]
 b. *There$_1$ seems [$_{IP}$ someone$_2$ to be [$_{PP}$ t$_2$ in the room]]

Given the predicate-internal subject hypothesis, *someone* is first merged in the SpecP position. When the derivation reaches the embedded IP domain, the Extended Projection Principle (EPP) becomes active; this local constraint requires filling of SpecI by either Merge or Move. Assuming that *there* is part of the numeration at this stage of the derivation, two possibilities arise: Either *there* is merged in SpecI (and subsequently raised to the matrix SpecI position), as in (56-a), or *someone* is moved to SpecI (and *there* is merged later in the derivation, directly in the matrix SpecI position), as in (56-b).

These two derivations involve an identical numeration, and they both respect all local constraints. In this case, Merge before Move tells us to choose the derivation underlying (56-b) and dispense with the derivation that generates (56-a).

Given that identity of numeration is a prerequisite for competition, (57) is correctly predicted to be possible – if there is no *there* present in the numeration, there is no competing derivation here that could be preferred by Merge before Move.

(57)   Someone$_1$ seems [$_{IP}$ t$'_1$ to be t$_1$ in the room]

The question arises of whether there is a deeper reason why Merge operations count as more economical than Move operations. Chomsky (1995, 1998) suggests that Move is to be defined in terms of Merge, which would make it inherently more complex, and this fact might ultimately be exploited in an attempt to derive the Merge before Move condition. Chomsky (1998:14) himself remarks: "Good design conditions would lead us to expect that simpler operations are preferred to more complex ones, so that Merge ... preempt[s] Move, which is a 'last resort,' chosen when nothing else is possible."

3.5 Conclusion

The four translocal constraints discussed so far do not yet exhaust the list of translocal constraints that have been proposed; see, e.g., the translocal Economy of Representation constraint in Chomsky (1991), or the translocal Preference Principle for Reconstruction in Chomsky (1993). Still, the constraints discussed here can be considered representative. At this point, we can address the question of what the structure of a minimalist syntax with translocal constraints looks like. Such a syntax has two parts. In the first part, derivations are created by structure-building (Merge), movement (Move), deletion, and perhaps other operations. Convergent derivations are assembled in reference sets according to criteria that must be decided on (see the above definitions of reference sets for some options). In the second part, translocal constraints choose among the competing derivations and thus determine the wellformedness of sentences. In essence, then, it turns out that a minimalist syntax with translocal constraints has exactly the shape that Prince & Smolensky (1993) attribute to an optimality-theoretic grammar: A first generator part (called *Gen*) creates the candidate set (= reference set, in minimalist syntax); Gen has only local constraints. A second "harmony"-evaluation part (called *H-Eval*)

determines the optimal candidate(s) (= derivation(s), in minimalist syntax) in the candidate set. More generally, we will see that all kinds of competition-based syntax have this structure, which is schematically shown in (58).

(58)   *Structure of a competition-based syntax*:
    a.   Gen creates the candidate set $\{C_1, C_2, ...\}$.
    b.   H-Eval determines the optimal candidate(s) C_i (C_j, ...) in
        $\{C_1, C_2, ...\}$.

Two issues concerning the notions of optimality and grammaticality in minimalist syntax remain to be clarified.[20] First, does optimality equal grammaticality? Whereas this is the case in optimality-theoretic syntax (see below), things are slightly more involved in minimalist syntax. As we have seen, it has been argued that derivations that converge can enter the competition, even though they may violate certain local constraints (recall the discussion of freezing effects in English). Accordingly, an optimal candidate may be one that violates a local constraint, and is therefore ungrammatical.

   Second, we have so far left open the question of how optimality evaluation proceeds in the presence of more than one translocal economy constraint in the grammar. In this case, conflicts may arise. As a simple, abstract example, suppose that there are two translocal constraints (T_1 and T_2), and only three derivations (D_1, D_2, and D_3) in the reference set.[21] Suppose further that T_1 prefers D_1 over D_2 and D_3; that T_2 prefers D_2 over D_1 and D_3; and that a derivation D_0 that would be preferred by both T_1 and T_2 fails to converge, so that it cannot participate in the competition. In such a situation, various possibilities arise. A first possibility would be what we can call "tolerance." On this view, it suffices to be selected by one translocal constraint to be optimal (hence, potentially grammatical); consequently, both D_1 and D_2 would be classified as optimal. A second possibility would be "ranking": The conflict among translocal constraints is resolved by a ranking, such that that derivation is optimal that is preferred by the higher-ranked constraint in the case of conflict. If T_1 is ranked higher than T_2, this would imply that only D_1 is optimal. Finally, a third possibility is what we can call "breakdown": In the case of conflicting instructions made by translocal constraints, no derivation can emerge as optimal. It turns out that this last possibility is the one that is generally assumed, and though it is not easy to come up with decisive evidence, it also strikes us as the most adequate one (see Collins 1994, Sternefeld 1997, and Müller 2000). On this basis, we can conclude that grammaticality can be defined as follows in a minimalist syntax with translocal constraints:

(59)   *Grammaticality*:
        A derivation D_i is grammatical iff (a) and (b) hold:
   a. D_i does not violate a local constraint.
   b. D_i is optimal.

(60)   *Optimality in minimalist syntax*:
        A derivation D_i is optimal iff there is no derivation D_j in the same
        reference set that is preferred over D_i by a translocal constraint.

The minimalist system that emerges in this way is not without problems. Some of those show up in all versions of competition-based syntax. For one thing, since a minimalist syntax of this type involves a global competition in a reference set that may be large, or even infinite, the overall complexity of the system is significantly increased. For another, we have seen that it is difficult to come up with a single, unified definition of reference set that accommodates all available evidence that one may want to treat in terms of translocal constraints.

  Other problems are more specific and confined to the particular notion of optimality that is employed in minimalist syntax. Most notably, the H-Eval metric is not maximally homogeneous and simple (because it may depend on a number of formally unrelated translocal constraints); however, it is rather inflexible nevertheless. Specifically, all translocal constraints must be classifiable as economy constraints in some sense (thus, properties of sentences that are not related to economy considerations cannot be subject to optimization). Even more importantly, it implies that all variation among languages must take place in the Gen part of the grammar – there is no room for parameterization in the H-Eval system. It is not always obvious that this position can be maintained in the light of conflicting empirical evidence. As an example, consider the effect that the Shortest Paths condition has on *wh*-movement in German. Recall that the Shortest Paths condition accounts for the superiority effect with *wh*-movement in languages like English; cf. (42). As has often been noted (see, e.g., Haider 1983), German does not exhibit superiority effects of this kind:

(61) a.   Ich frage mich   [$_{CP}$ wer$_1$ C t$_1$ was$_2$ gekauft hat]
          I   ask   myself   who   what bought has
     b.   Ich frage mich   [$_{CP}$ was$_2$ C wer$_1$ t$_2$ gekauft hat]
          I   ask   myself   what   who   bought has

Still, it seems clear that the path from t$_1$ to *wer*$_1$ in (61-a) is shorter than the path from t$_2$ to *was*$_2$ in (61-b). To avoid the result that the Shortest Paths

condition blocks (61-b) in favor of (61-a), additional assumptions concerning Gen are therefore necessary.[22]

Not least because of problems like this, there is a strong tendency in recent versions of the minimalist program to dispense with translocal constraints – and hence, with the concept of competition – altogether; see in particular Collins (1997), Frampton & Gutman (1999), and also Chomsky (1995, 1998).[23] Local (derivational) constraints like the Last Resort condition and the Minimal Link Condition (MLC) have been developed in Chomsky (1995) and much recent work as economy conditions that can take over at least some of the work that was done by translocal constraints like Fewest Steps and Shortest Paths. Similarly, effects that were attributed to Procrastinate and Merge before Move have been shown to be derivable without invoking translocal constraints.

Apart from these theory-internal considerations, it is interesting to note that the fall of translocal constraints (and with it the fall of the concept of competition) in minimalist syntax goes hand in hand with the rise of optimality theory, and hence optimality-theoretic syntax, which inherently relies on translocality and competition. However, before turning to this approach, we will discuss another model of competition-based syntax, one that developed concurrently with (and as an extension of) government and binding theory: blocking syntax.

4   Blocking Syntax

Blocking syntax was developed by DiSciullo & Williams (1987) and Williams (1997) on the basis of Aronoff's (1976) approach to blocking in morphology. To a large extent, it is equivalent to a syntactic theory that incorporates the Elsewhere Condition which has played a major role in phonology (see Kiparsky 1982 and the references cited there). A version of blocking syntax that is even closer to Kiparsky's approach is developed in Fanselow (1989, 1991); this approach relies on the Proper Inclusion Principle. A blocking syntax has the same general form as a minimalist syntax, viz., that in (58): Gen generates the candidates (in blocking syntax typically S-structure representations) which are assembled in candidate sets. The competing candidates are then subjected to an H-Eval procedure that determines the optimal candidate(s).

The underlying idea of blocking syntax is that synonyms are not tolerated

in natural languages. Consequently, candidate sets are defined in terms of identity of meaning:

(62)   *Candidate Set*:
       Two candidates C_i and C_j are in the same candidate set iff they (i) have the same meaning, and (ii) respect all local constraints.

A candidate is grammatical iff it is optimal (given (62), an optimal candidate cannot violate a local constraint). The concept of optimality is different from that adopted in the minimalist program. The optimal candidate is the most specific one. Thus, more specific candidates block less specific ones: This is the Blocking Principle.

(63)   *Optimality in blocking syntax*:
       A candidate C_i is optimal iff there is no candidate C_j in the same candidate set that is more specific.

This approach crucially depends on how specificity is understood. In morphology, irregular forms count as more specific than regular forms. In syntax, we can assume that C_i is more specific than C_j if local constraints (or rules) lead to C_i's distribution being more restricted than that of C_j.

4.1   Comparative Formation in Williams (1997)

With this in mind, let us consider an example that is discussed in Williams (1997): English comparative formation.

   As shown in (64), English has two ways of comparative formation: a morphological one and a syntactic one. The two strategies appear to be in complementary distribution.

(64) a.   hot → hotter, *more hot
     b.   happy → happier, *more happy
     c.   colorful → *colorfuller, more colorful

Williams suggests that these data can be accounted for by the following two rules, which we call rule A and rule B.

(65) a.   *Rule A* (morphological):
          Comparatives can be formed by attaching the suffix *er* to monosyllabic adjectives, and to disyllabic adjectives ending in *y*.

b.   *Rule B* (syntactic):
Comparatives can be formed by adding *more* in the syntax.

The candidate that employs the morphological comparative **colorfuller* violates rule A, while the candidate that uses the syntactic comparative *more colorful* respects both rule A (vacuously) and rule B. In contrast, candidates like *hotter* or *happier* respect rule A (and, vacuously, rule B); but what is left open by rule A and rule B is why candidates like **more hot* or **more happy* are ungrammatical. The illformedness of these forms could be derived by replacing rule B with rule B′ in (66). However, adopting this rule would lead to a redundancy: The context that permits morphological comparatives in rule A is repeated in an identical form as the context that prohibits syntactic comparatives in rule B′.

(66)   *Rule B′*:
Comparatives can be formed by adding *more* in the syntax, unless the adjective is monosyllabic, or disyllabic and ending in *y*.

To avoid this redundancy, Williams proposes maintaining rule B instead of adopting rule B′. The generalization that syntactic comparatives seem to be possible only if morphological comparatives are impossible is then derived by invoking the notion of blocking embodied in (63). Given rules A and B, morphological comparatives are more restricted in their distribution than syntactic comparatives – in fact, rule B imposes no specific restrictions on syntactic comparative formation at all. Hence, if both morphological and syntactic comparatives respect all local constraints (like rules A and B), the more specific morphological comparative is selected as optimal. If, however, the morphological comparative violates a local constraint (like rule A), the syntactic comparative cannot be blocked anymore, and is consequently selected as optimal.

4.2   Anaphors vs. Pronouns in Fanselow (1991)

Recall from section 2.1 the data that seem to suggest a complementary distribution of anaphors and pronominals in English (at least in the domain under discussion here).

(67) a.   John$_1$ likes himself$_1$
b.   *John$_1$ thinks that Mary likes himself$_1$
c.   John$_1$ thinks that Mary likes him$_1$

     d. *John$_1$ likes him$_1$

We have seen that standard government and binding theory accounts for these data by invoking the principles A and B in (68).

(68) a. *Principle A*:
        An anaphor is bound in its binding domain.
     b. *Principle B*:
        A pronominal is free in its binding domain.

However, as in the case of the comparative formation rules A and B$'$ that were just discussed, it seems that this approach involves a redundancy: A generalization is missed if two separate local constraints are postulated for anaphors and pronominals, where the context that permits one strategy is identical to the context that precludes the other strategy (viz., the binding domain in both cases). As noted by Fanselow (1989, 1991), Burzio (1991), and Richards (1997), among others, a more elegant account can be given if the notion of competition is invoked. Here, we will sketch Fanselow's blocking approach.[24]

    Fanselow's analysis relies on the Proper Inclusion Principle (PIP), a version of the Elswhere Condition (cf. Kiparsky 1982) that can be viewed as a translocal constraint:

(69)   *Proper Inclusion Principle* (PIP):
     a. Suppose that two feature assignment mechanisms M_1 and M_2 compete in a given structure. Then, other things being equal, M_2 cannot be applied if M_1 is more specific.
     b. M_1 is more specific than M_2 if the application domain of M_2 properly includes the application domain of M_1.

The feature assignment mechanisms that play a role in the present context are (a) the assignment of the feature [+anaphoric] to an NP, and (b) the assignment of the feature [+pronominal] to an NP – in short, reflexivization (or reciprocalization) and pronominalization. By assumption, the assignment of the feature [+anaphoric] is subject to (something like) Principle A, whereas there is no comparable requirement for the assignment of the feature [+pronominal]; i.e., Principle B is dropped. This implies that, due to Principle A, anaphors are more restricted in their distribution than pronominals; the application domain of pronominalization properly includes the application domain of reflexivization. From this it follows directly that in all those cases where both anaphors and pronominals respect all local constraints, the PIP

forces the choice of the anaphor. Pronominals can emerge only in contexts in which anaphors are precluded (e.g., because of a violation of Principle A, as in the examples presently under consideration).

The PIP can be viewed as a version of the blocking principle that is part of the definition of optimality in (63). The only relevant change that must be made for the case at hand concerns the question of which entities compete. We can now assume that the competing items are complete syntactic objects (syntactic candidates), rather than feature assignment mechanisms.

4.3   Conclusion

Blocking syntax is characterized by the fact that it is fairly simple in various respects. Most importantly, blocking syntax employs a simple concept of optimality in its H-Eval part. There is only one translocal constraint (the Blocking Principle that selects the most specific candidate), not more than one, as in minimalist syntax. In addition, due to the origin of the blocking principle as a means to avoid synonymy, blocking analyses uniformly rely on identity of meaning in the definition of candidate sets, again in contrast to the variability involved in minimalist syntax. However, the simplicity comes at a certain price: Harmony evaluation is even less flexible than in minimalist syntax. First, the blocking principle by its very nature can only have a small domain in which it is active; in general, the role of H-Eval in optimality theory is smaller than in the minimalist program. Second, there is no room for parameterization in the H-Eval domain at all. And third, since blocking analyses depend on complementarity of distribution, cases of optionality pose problems that are almost insurmountable.[25] A competition-based approach that strengthens the role of the H-Eval part of the grammar and increases flexibility in this domain is optimality-theoretic syntax; and it is this model that we finally turn to now.

5   Optimality-Theoretic Syntax

5.1   Basic Concepts

By definition, an optimality-theoretic syntax takes the general form in (58), with the grammar divided into a Gen part that creates the competing candidates, and an H-Eval part that selects the optimal candidate(s). Recall that the notion of optimality in a minimalist syntax or in a blocking syntax is a

comparatively simple one: Optimality is determined by a small set of simple translocal economy constraints in the former case (cf. (60)), and by a single translocal blocking principle (selecting the most specific candidate) in the latter (cf. (63)). In optimality-theoretic syntax, there is only one translocal constraint that determines optimality: Optimal (and grammatical) is a candidate that has the best "constraint profile" – or, more precisely, a candidate for which there is no competitor that has a better constraint profile; cf. (70). This definition makes it possible for more than one candidate to be optimal in a given candidate set.

(70)   *Optimality in optimality-theoretic syntax*:
         A candidate C_i is optimal (= grammatical) iff there is no candidate C_j in the same candidate set that has a better constraint profile.

However, the evaluation metric is internally highly complex. The notion of constraint profile is defined in (71).

(71)   *Constraint Profile*:
         A candidate C_j has a better constraint profile than a candidate C_i iff there is a constraint Con such that (a) and (b) hold:
   a. C_j satisfies Con better than C_i; i.e., C_j satisfies Con and C_i violates Con, or C_j violates Con less often than C_i.
   b.   There is no constraint Con′ ranked higher than Con on which C_i and C_j differ.

This presupposes that in addition to the local constraints employed by the Gen component, which are inviolable and unranked, the H-Eval component relies on a system of local constraints that are violable and ranked (and, by assumption, universal) in order to determine the best constraint profile, hence, optimality. The ranking among the violable local constraints of the H-Eval component is indicated by the symbol $\gg$; the H-Eval constraints themselves are typically written with small capitals. Optimality-theoretic competitions are often illustrated by tables (so-called *tableaux*); optimality of a candidate is indicated by the *pointing finger*: ☞ ; violation of a local constraint is shown by a star * in the appropriate column of the table; if this violation is fatal for a candidate (i.e., responsible for its suboptimality), an exclamation mark ! is added (redundantly). In the abstract H-Eval competition in table T_1, in which the candidate set consists of C_1-C_5, C_1 emerges as the optimal candidate: It avoids a violation of the high-ranked constraints A and B (unlike C_3-C_5), and

it minimizes a violation of the low-ranked constraint C (unlike C_2). Hence, there is no competing candidate with a better constraint profile than C_1.

T_1: Determining optimality

Candidates	A	B	C
☞ C_1			*
C_2			**!
C_3		*!	
C_4	*!		
C_5		*!	*

By reranking the constraints B and C in T_1, candidate C_3 emerges as the optimal candidate; cf. table T_2.

T_2: Reranking

Candidates	A	C	B
C_1		*!	
C_2		*!*	
☞ C_3			*
C_4	*!		
C_5		*!	*

Reranking of constraints forms the basis of the concept of parameterization in optimality-theoretic syntax. A further characteristic feature of this approach is that it is essentially non-cumulative; i.e., no number of violations of a low-ranked constraint can outweigh a single violation of a higher-ranked constraint. Thus, suppose that there were an additional, lowest-ranked constraint D in T_1 that C_1 violated, say, five times, and that C_2-C_5 did not violate at all. This would not undermine C_1's optimality.

Before we turn to some illustrations of optimality-theoretic analyses, something must be said about the nature of candidates and candidate sets. Optimality-theoretic syntax is strongly influenced by work in optimality-theoretic phonology. Since the latter is characterized by an orientation that is predominantly representational (cf. Prince & Smolensky 1993 and McCarthy & Prince 1995), it does not come as a surprise that many approaches in optimality-theoretic syntax postulate that the competing candidates created by Gen are surface structure representations. This holds, e.g., for what can arguably be viewed as the three most influential analyses in optimality-theoretic syntax so far, viz., Grimshaw (1997), Pesetsky (1998), and Legendre, Smolensky & Wilson (1998). However, there is no inherent reason

why the candidates that are subject to optimization should not be syntactic objects of a more complex type, like $\langle$D-structure, S-structure, LF$\rangle$ tuples as in government and binding theory, or, indeed, complete derivations, as in the minimalist program.[26] The choice of candidate type goes hand in hand with the choice of local constraint type that shows up in the H-Eval part as violable and ranked: If candidates are representations, constraints will be representational; if candidates are derivations, constraints will be derivational; and if candidates are $\langle$D-structure, S-structure, LF$\rangle$ tuples as in government and binding theory, constraints can take any of the forms sketched in section 2.

Similarly, candidate sets can be defined in various ways, which of course significantly influences the nature of the competition. Basically, all of the definitions of reference sets in minimalist syntax that have been proposed (see section 3 and Sternefeld 1997) are also potential definitions of candidate sets in optimality-theoretic syntax. A further influential definition of candidate sets comes from Grimshaw (1997). She postulates that two candidates (S-structure representations) compete iff they are realizations of the same predicate/argument structure and have non-distinct logical forms (or non-distinct interpretations).

By making optimality depend on an intricate system of violable and ranked constraints, H-Eval – and hence, the concept of competition – becomes even more important than in minimalist syntax and blocking syntax. As a matter of fact, much work in optimality-theoretic syntax has tried to minimize the role of the Gen component, and maximize the role of the H-Eval component (but see Pesetsky 1997, 1998 for some cautionary remarks).

An optimality-theoretic approach gains immediate support in all those contexts where postulating a competition of syntactic objects is initially plausible. This includes, but is by no means confined to, contexts where notions of economy seem to play a role. A prototypical case is one in which the wellformedness of a sentence S_i that exhibits an otherwise peculiar property seems to depend on the unavailability of another sentence S_j that exhibits the property one would normally expect. Here, S_i is often referred to as a "repair" form; a typical instance is the English *do*-support construction. Accordingly, *do*-support was among the first phenomena to be tackled in optimality-theoretic syntax (see Speas 1995 and Grimshaw 1997). Most of the constructions discussed in sections 2-4 can also be viewed as suggesting an underlying competition; and indeed, they can fruitfully be addressed in optimality-theoretic syntax. This is shown in the following section.

5.2 Case Studies

5.2.1 *Anaphors vs. Pronouns in Wilson (1999)*

Let us begin with the competition between reflexivization and pronominalization. The following optimality-theoretic account is based on Wilson (1999).[27] Recall the generalization that, by and large, pronominals are allowed to express binding relations in English in just those cases in which anaphors are not allowed to do so. To account for this, a prerequisite is that two sentences which differ only with respect to the choice of anaphor vs. pronominal in a given position must compete. The ranking LOC-ANT $\gg$ REF-ECON of the two constraints in (72) then produces the right results.

(72) a. LOC-ANT ("Local Antecedent"):
          If a binding domain contains an anaphor, then it must also contain the anaphor's antecedent.
      b. REF-ECON ("Referential Economy"):
          A (referentially dependent) argument must not have lexical ϕ-feature specification.

LOC-ANT is a version of Principle A; it requires a local antecedent for an anaphor. Hence, as a tendency, this constraint favors pronominals, which satisfy it vacuously. REF-ECON, on the other hand, inherently prefers anaphors to pronominals if we are willing to make the assumption that anaphors do not have a lexical ϕ-feature specification, whereas pronominals do. Consequently, when an anaphor can respect LOC-ANT, the violation of REF-ECON incurred by a pronominal is fatal; cf. table T_3.

T_3: Reflexivization

Candidates	LOC-ANT	REF-ECON
☞ C_1: John$_1$ likes himself$_1$		
C_2: John$_1$ likes him$_1$		*!

However, when an anaphor cannot find an antecedent in its binding domain and must violate LOC-ANT, a violation of the lower-ranked REF-ECON constraint becomes possible, and pronominalization turns out to be optimal; cf. table T_4.

T$_4$: Pronominalization

Candidates	LOC-ANT	REF-ECON
C$_1$: John$_1$ thinks that Mary likes himself$_1$	*!	
☞ C$_2$: John$_1$ thinks that Mary likes him$_1$		*

5.2.2   *Complementizer-Trace Effects in Grimshaw (1997)*

In section 2, we noted that government and binding theory accounts for the complementizer-trace effect in (4-a) on a purely local basis, without postulating a competition with the complementizer-less variant in (4-b) from which only the latter would emerge as optimal. This view is abandoned in Déprez (1991), which is the basis of the optimality-theoretic account advanced in Grimshaw (1997). As background, Grimshaw assumes that the size of clauses is variable. Clauses are extended projections of V; they are minimally VPs, but they can be IPs, CPs, or functional projections of an even bigger size, depending on the outcome of optimization. Bridge verbs in English permit both CP-embedding (with a complementizer – a declarative CP without a complementizer will typically fatally violate a high-ranked constraint that precludes empty head positions) and IP- or VP-embedding (without a complementizer). In the latter case, IP must be chosen if an auxiliary or *do* is present (i.e., if the need arises to accommodate an additional lexical head); VP can be chosen otherwise. The main constraints that are needed in the account of complementizer-trace effects are listed in (73). A possible ranking for English is OP-SPEC ≫ T-LEX-GOV ≫ STAY.[28]

(73) a.   OP-SPEC ("Operator in Specifier"):
         Wh-operators must occupy a specifier position from which they c-command all elements of the extended V projection over which they take scope.
      b.   T-LEX-GOV ("Lexical Government of Traces"):
         A trace is lexically governed.
      c.   STAY ("Economy of Movement"):
         Trace is not allowed.

STAY is a local version of the translocal economy constraint Fewest Steps; OP-SPEC is a version of the *Wh*-Criterion that is often postulated in government and binding theory (see, e.g., Lasnik & Saito 1992). The ranking OP-SPEC ≫ STAY ensures overt *wh*-movement in simple questions in English.

T-LEX-GOV corresponds to (a part of) the ECP. Assuming that candidates with and without *that* compete, the complementizer-trace effect is derived as shown in table T_5.[29]

T_5: *Subject wh-movement*

Candidates	OP-SPEC	T-LEX-GOV	STAY
C_1: ... who$_1$ you think [$_{CP}$ that [$_{IP}$ t$_1$ will leave]]		*!	*
☞ C_2: ... who$_1$ you think [$_{IP}$ t$_1$ will leave]			*
C_3: ... you think [$_{CP}$ that [$_{IP}$ who$_1$ will leave]]	*!		
C_4: ... you think [$_{IP}$ who$_1$ will leave]	*!		

C_3 and C_4 fatally violate OP-SPEC. Both C_1 and C_2 violate STAY, but C_1 violates T-LEX-GOV in addition: t_1 in C_1 is not lexically governed (*that* being unable to do so), whereas t_1 in C_2 is lexically governed (by the matrix V).[30] In contrast, an embedded V governs object traces throughout, irrespective of the presence or absence of a complementizer *that*; hence, T-LEX-GOV is satisfied equally well by C_1 and C_2 in table T_6. Given that C_1 and C_2 do not differ with respect to any other constraint either, optionality of a complementizer is correctly predicted in cases of object extraction, due to an identical constraint profile.

T_6: *Object wh-movement*

Candidates	OP-SPEC	T-LEX-GOV	STAY
☞ C_1: ... who$_1$ you think [$_{CP}$ that [$_{IP}$ she will invite t$_1$]]			*
☞ C_2: ... who$_1$ you think [$_{IP}$ she will invite t$_1$]			*
C_3: ... you think [$_{CP}$ that [$_{IP}$ she will invite who$_1$]]	*!		
C_4: ... you think [$_I$ she will invite who$_1$]	*!		

Thus far, there is no evidence for treating T-LEX-GOV as a violable constraint in the H-Eval part of the grammar (rather than as an inviolable constraint in the Gen part). Such evidence can be gained by considering adjunct extraction. In this case, T-LEX-GOV is violated by both candidates involving *wh*-movement (C_1, C_2). However, given that there is no competing candidate that can avoid a violation of T-LEX-GOV without violating a higher-ranked constraint (e.g., a candidate that employs a resumptive pronoun; see below), C_1 and C_2 can emerge as optimal despite this violation.

T₇: Adjunct wh-movement

Candidates	Op-Spec	T-Lex-Gov	Stay
☞ C₁: ... why₁ you think [CP that [IP she has left t₁]]		*	*
☞ C₂: ... why₁ you think [IP she has left t₁]		*	*
C₃: ... you think [CP that [IP she has left why₁]]	*!		
C₄: ... you think [IP she has left why₁]	*!		

5.2.3 *Subjacency and Resumptive Pronouns in Pesetsky (1998) and Legendre, Smolensky & Wilson (1998)*

Recall that resumptive pronouns often seem to be possible only as last resort strategies in cases where traces are blocked (see (8)). Competition-free models like government and binding theory have no obvious means to relate one construction to the other (at least, as long as they are supposed to stay strictly competetion-free; see note 3); but the case is different in optimality-theoretic syntax. An optimality-theoretic account of resumptive pronoun strategies is developed in Legendre, Smolensky & Wilson (1998) (on the basis of evidence from Chinese) and Pesetsky (1998) (on the basis of English data comparable to those in (8), as well as evidence from Hebrew, Russian, and Polish). The details of the two analyses differ a great deal, but the gist of the explanation is identical; it centers around two constraints like those in (74).[31]

(74) a. CNPC ("Complex Noun Phrase Condition"):
Traces must not be separated by a complex noun phrase from their antecedents.
 b. RES ("Resumptive Constraint"):
Resumptive pronouns are prohibited.

The CNPC prohibits traces in certain (non-local) environments; RES disfavors resumptive pronouns (i.e., pronouns that are bound from an A-bar position) in general. The ranking is CNPC ≫ RES. (Thus, the two constraints and their ranking are analogous to what we have seen with LOC-ANT and REF-ECON in the domain of binding theory.) As with the *wh*-movement construction discussed in the last section, it must be ensured that overt movement of the relative operator takes place in examples like those in (8). We assume that this is independently taken care of.[32] Based on these assumptions, consider table T₈.

T_8: Trace vs. resumptive pronoun in transparent contexts

Candidates	CNPC	RES
☞ C_1: the man [$_{CP}$ who(m)$_1$ I saw t$_1$]		
C_2: the man [$_{CP}$ who(m)$_1$ I saw him$_1$]		*!

Both candidates respect CNPC. Consequently, the RES violation incurred by the resumptive pronoun in C_2 becomes fatal, and C_1 is optimal. However, in the competition illustrated in table T_9, C_1 violates CNPC. In this case, C_2's RES violation is tolerable, and the resumptive pronoun strategy emerges as optimal.[33]

T_9: Trace vs. resumptive pronoun in CNPC *contexts*

Candidates	CNPC	RES
C_1: the man [$_{CP}$ who(m)$_1$ [$_{IP}$ I don't believe [$_{NP}$ the claim [$_{CP}$ t$'_1$ that anyone saw t$_1$]]]]	*!	
☞ C_2: the man [$_{CP}$ who(m)$_1$ [$_{IP}$ I don't believe [$_{NP}$ the claim [$_{CP}$ that anyone saw him$_1$]]]]		*

A lot more could be said about relativization in English and other languages in an optimality-theoretic approach (in particular, concerning *that*-relatives and their relation to *wh*-relatives), but these considerations will have to suffice for now; cf. Grimshaw (1997) and Pesetsky (1998).

5.2.4 *Avoid Pronoun*

Consider now the Avoid Pronoun facts that were discussed in section 2 (cf. (10) and (12)). In English gerunds, PRO and a lexical pronoun can both occur in principle; however, PRO must be used instead of a lexical pronoun if it can fulfill the Control Rule. A transfer of Chomsky's (1981) approach into optimality theory is straightforward. The Control Rule in (11) can directly be viewed as an optimality-theoretic constraint (with the same qualification as in government and binding theory; see note 4); cf. (75-a). The Avoid Pronoun Principle in (13) can be simplified by turning this translocal constraint into a local (though violable) one; cf. (75-b).[34] The ranking for English is CONTROL ≫ *PRON.

(75) a.   CONTROL ("Control Rule"):
        If PRO is minimally dominated by a declarative clausal object α,

then it must be controlled by an antecedent within the minimal CP
dominating α.

b. *PRON ("Avoid Pronoun"):
Pronouns are prohibited.

Suppose that candidate sets are defined in such a way that candidates with
PRO and candidates with a lexical pronoun can compete, but, crucially, that
sentences with different indexings (hence, different logical forms) do not
compete. Then, the facts fall into place. The blocking of a lexical pronoun
by PRO in cases where CONTROL can be satisfied is illustrated in table T_{10}.

T_{10}: PRO vs. pronoun under co-indexing

Candidates	CONTROL	*PRON
C_1: John$_1$ would much prefer [his$_1$ going to the movie]		*!
☞ C_2: John$_1$ would much prefer [PRO$_1$ going to the movie]		

Table T_{11} illustrates the case where PRO is not co-indexed with the matrix
antecedent, thereby violating CONTROL. Here, the *PRON violation incurred
by all pronouns is non-fatal, and the pronoun strategy is optimal.

T_{11}: PRO vs. pronoun under contra-indexing

Candidates	CONTROL	*PRON
☞ C_1: John$_1$ would much prefer [his$_2$ going to the movie]		*
C_2: John$_1$ would much prefer [PRO$_2$ going to the movie]	*!	

5.2.5 *Superiority Effects*

The question arises of whether the evidence that is accounted for by translo-
cal constraints in minimalist syntax can also be reanalyzed in optimality-
theoretic syntax. At least to some extent, this seems to be the case. As
noted above, the STAY constraint adopted in Grimshaw (1997), Legendre,
Smolensky & Wilson (1998), and much related work, is essentially a local
version of the translocal Fewest Steps condition. Similarly, a local counter-
part has been suggested for the translocal economy constraint Shortest Paths.
Let us reconsider the superiority phenomenon as one of the core applications
of Shortest Paths. Some relevant examples are repeated in (76).

(76) a.   I wonder [$_{CP}$ who$_1$ C [$_{IP}$ t$_1$ bought what$_2$]]
    b. *I wonder [$_{CP}$ what$_2$ C [$_{IP}$ who$_1$ bought t$_2$]]

Ackema & Neeleman (1998) propose a local version of Shortest Paths that we may call Min-Chain ("Minimize Chain Length"). This constraint records a star * for every node crossed by a movement chain.[35] Assuming that only overt movement counts for the purposes of this constraint, (76-a) can successfully block (76-b) under Min-Chain: Other things being equal, the *wh*-chain in (76-a) violates Min-Chain twice (IP and C′ are crossed), whereas the *wh*-chain in (76-b) violates Min-Chain four times (VP, I′, IP, and C′ are crossed), the third violation being fatal already.

Another account is developed by Legendre, Smolensky & Wilson (1998). They start out with Bar, which is nearly identical to the CED given above:

(77)   Bar ("Barriers Condition"):
         A chain link must not cross a barrier.

By a general operation of *local constraint conjunction*, Bar can be conjoined with various constraints, including itself. Reflexive local conjunction of Bar and Bar yields a new constraint Bar & Bar = Bar^2 that is violated if a chain link crosses two barriers. Local conjunction of Bar^2 and Bar yields Bar^3, which is violated if a chain link crosses three barriers, and so on. This mechanism recursively produces a subhierarchy of Bar^i constraints which has a fixed internal ranking, given a universal meta-restriction on constraint conjunction: Con_1 & $\text{Con}_2 \gg \text{Con}_1, \text{Con}_2$.[36] The violability of Bar^i subhierarchy constraints makes it possible to adopt a simple theory of barriers according to which a significant number of XPs are barriers (see Koster 1987 for this general idea). Adopting a strict interpretation of Chomsky (1986a), Legendre, Smolensky & Wilson (1998) assume that all non-L-marked XPs are barriers, including VP, IP, subject NPs, adjunct CPs, etc. This makes the Bar^i subhierarchy a good means to measure path lengths and derive typical Shortest Paths effects. In the superiority case currently under consideration, subject *wh*-movement as in (76-a) crosses only an IP barrier, violating Bar, whereas object *wh*-movement as in (76-b) crosses a VP barrier and an IP barrier, violating Bar^2, which is universally ranked higher. Given that the two candidates do not differ with respect to other constraints, and that they have a better constraint profile than all other competitors, it follows that the availability of short subject movement blocks the possibility of longer object movement, as desired.

To end this section, note that Legendre, Smolensky & Wilson (1998) also succeed in reanalyzing the Tagalog *wh*-movement evidence discussed above in an optimality-theoretic way by invoking the Bar^i subhierarchy. More generally, we can conclude that all of the analyses involving translocal constraints

that have been proposed in minimalist syntax or blocking syntax can be recast in optimality-theoretic terms by employing local, violable constraints.[37]

5.3 Some Open Issues

Optimality-theoretic syntax inherits the complexity problem from minimalist syntax. Candidate sets are typically large (as in Pesetsky 1998), often infinite (as in Grimshaw 1997). In addition, there are several open issues that are specific to the optimality-theoretic approach. We will focus on two of these in what follows, and briefly mention five others after that.

5.3.1 *Inputs and Faithfulness*

An important optimality-theoretic concept that has played no role so far in our discussion is the notion of *input*. In optimality theory (cf. Prince & Smolensky 1993), Gen does not create competing candidates (the *outputs*) freely; rather, it does so on the basis of a given input. In phonology, inputs are underlying representations stored in the lexicon; here, inputs qualify as roughly the same types of objects as outputs. In syntax, it is much less clear what the input might look like (see Archangeli & Langendoen 1996). The null hypothesis – that the input is a completely articulated potential sentence of the same type as the output candidates – is not unproblematic because it would seem to imply the assumption that all possible sentences are "stored," which cannot possibly be true. To find out what the input in syntax is, it is instructive to consider its theory-internal functions. By and large, there are two. First, the input is standardly taken to define the competition. Second, the input serves as a basis for *faithfulness* constraints that demand input/output identity and thereby minimize deviations from the input in the optimal output. Let us consider the second function first.

Faithfulness constraints play an important role in phonology. Constraints of the PARSE (or MAX) family prohibit deletion of input material in the output; constraints of the FILL (or DEP) family prohibit insertion of output material that is not part of the input; constraints of the IDENT family prohibit modifying input material. Faithfulness constraints have also been adopted in much recent work in optimality-theoretic syntax. The following two constraints are taken from Legendre, Smolensky & Wilson (1998) and Baković & Keer (1999), respectively.

(78) a.   PARSE[SCOPE]:
            Scope assignment in the input must be realized by chain formation
            in the output.
     b.   FAITH[COMP]:
            The output value of [±COMP] is the same as the input value.

Note that (78-a) implies that the input is a more complex object than just a
collection of words (a numeration) or a predicate/argument structure; it must
be a highly structured representation that encodes the relative scope of op-
erators. (78-b) presupposes an abstract feature [±COMP] that for the present
purposes we can assume to be located on a V that selects a proposition. Let
us consider candidates that violate these constraints. Suppose that (79-a) is
the input for output candidate (79-b), and (79-c) is the input for output candi-
date (79-d). Legendre, Smolensky & Wilson (1998) assume that (79-b) vio-
lates PARSE[SCOPE] because matrix scope for *how*$_1$ in the input (79-a) (indi-
cated by [+wh]$_1$) is reduced to embedded scope in the output (again indicated
by [+wh]$_1$). Similarly, Baković & Keer (1999) assume that (79-d) violates
FAITH[COMP] because a [–COMP] specification in the input contrasts with a
[+COMP] specification (hence, a complementizer) in the output.[38]

(79) a.   [+wh]$_1$ ... wonder$_{[+wh]}$ [[+wh]$_2$ ... what$_2$ ... how$_1$ ...]   (input)
     b.   You wonder$_{[+wh]}$ [$_{CP}$ [+wh]$_1$ [+wh]$_2$ how$_1$ John did what$_2$] (output)
     c.   ... V$_{[-Comp]}$ [...]   (input)
     d.   I think [$_{CP}$ that [$_{PP}$ on him]$_1$ no coat looks good t$_1$]   (output)

At this point, we need not go into the actual analyses in which these con-
straints play a role (as it happens, both faithfulness violations turn out to
be non-fatal, i.e., (79-b,d) are optimal). The crucial question is: Is it really
necessary to refer to the concept of input here, or is it possible to read the
respective violations off the output forms, without any reference to inputs?
At least for the cases at hand, the answer is straightforward: By enriching
output representations in ways that have independently been proposed, a ref-
erence to inputs becomes unnecessary. (79-a,b) is a case where the intended
matrix scope is not reached by chain formation in the candidate. Employing
abstract scope markers (Σ) in S-structure representations (cf., e.g., Williams
1986), we can equivalently encode this input information in the output, as
in (80-a).[39] As for the case in (79-c,d), the only assumption that we have to
make (and which strikes us as innocuous, in fact, completely standard) is that
selectional properties of lexical heads are accessible in syntax; cf. (80-b).

(80) a. Σ_1 you wonder$_{[+wh]}$ [$_{CP}$ [+wh]$_1$ [+wh]$_2$ how$_1$ John did what$_2$]

                                                                 (output)

    b.   I think$_{[-Comp]}$ [$_{CP}$ that [$_{PP}$ on him]$_1$ no coat looks good t$_1$] (output)

PARSE[SCOPE] and FAITH[COMP] can now be modified in obvious ways, without reference to inputs.

(81) a.   PARSE[SCOPE] (revised):
         Scope markers must be reached by chain formation.
    b.   FAITH[COMP] (revised):
         Lexical [±COMP] selection requirements must be respected.

If this result can be generalized, and all syntactic faithfulness constraints can be reanalyzed in this way, we can conclude that these constraints do not support the concept of input anymore. Why should it be that the notion of input is relevant for phonological faithfulness constraints, but not for their syntactic counterparts? The answer, we believe, follows from what appears to be a fundamental difference between syntax and phonology: Syntax is an information-preserving system with richly structured output candidates, whereas phonology is a system that loses information, so that reference to an underlying input is necessary in constraints.

    With this in mind, let us turn to the other input function noted above, that of defining candidate sets. Since syntactic output candidates are richly structured, all the relevant information that they must share in order to compete can be read off them, independently of what notion of candidate set is adopted; again, this is in sharp distinction to phonology. Thus, it is possible to explicitly define candidate sets without reference to the concept of input. For instance, if we follow Grimshaw (1997) in assuming that competing candidates must have the same predicate/argument structure, we can read this information off the potentially competing candidates themselves.

    As a matter of fact, it turns out that an input-independent characterization of candidate sets cannot even be avoided in Grimshaw's own approach. Recall that Grimshaw (1997) postulates that two candidates compete only if they have non-distinct logical forms (in addition to identical predicate/argument structures). If the input fully determines the candidate set, this presupposes that an input is a complex object that exhibits all relevant logical form information. It is generally assumed that outputs can deviate from inputs in many ways, subject only to faithfulness constraints. Hence, if nothing else is said, we expect that output candidates can be semantically unfaithful to the input by, e.g., applying scope reduction (such that, e.g., a *wh*-phrase with

matrix scope in the input is interpreted with embedded scope in the output). This clearly implies that candidates with distinct logical forms can compete. This consequence is embraced by Legendre, Smolensky & Wilson (1998) (cf. (79-b)). However, such a result is incompatible with Grimshaw's (1997) assumptions, according to which competing candidates must *have* (not: *go back to*) non-distinct logical forms. Thus, even in this approach, the input cannot completely determine the competition; the requirement of non-distinct logical forms must be stipulated on top of it.

More generally, it emerges that an input-free characterization of candidate sets is both readily available and independently motivated. Hence, reference to inputs is unnecessary for the purpose of defining competition in syntax.

From all this, we would like to conclude that it may eventually be possible to dispense with the notion of input in syntax; but further research is needed in this domain (also see note 43 below).

5.3.2 *Absolute Ungrammaticality*

Another important open question in optimality-theoretic syntax is how to account for the phenomenon of "absolute ungrammaticality" or "ineffability," i.e., cases where there does not seem to be a candidate in a candidate set that is grammatical. As an example, consider the following ungrammatical example involving *wh*-extraction across an adjunct island in German:

(82)   *Was_1 ist Fritz eingeschlafen [$_{CP}$ nachdem er t_1 gelesen hat] ?
         what is Fritz fallen asleep   after   he   read   has

Let us apply the suggestions that can be found in the literature to the case at hand. First, Pesetsky (1997, 1998) emphasizes that certain sentences may be ungrammatical not because they are classified as suboptimal in the H-Eval part of the grammar, but because they cannot be generated by Gen in the first place. Thus, a constraint like (83) might be part of Gen.

(83)   ADJ-ISL ("Adjunct Island Constraint"):
         A trace must not be separated by an adjunct clause from its antecedent.

Second, it is suggested in Grimshaw (1994) and Müller (1997) that certain optimal candidates may have properties that make them inaccessible for other domains of the language faculty, like, e.g., semantic interpretation. ADJ-ISL might be part of H-Eval, but ranked higher than OP-SPEC. On this view,

(84) could block (83) as suboptimal; but this optimal candidate would be uninterpretable (indicated by #) and, hence, unusable.

(84)   #Fritz ist eingeschlafen [$_{CP}$ nachdem er was$_1$ gelesen hat] ?
           Fritz is fallen asleep   after   he what read   has

These two approaches have in common that they allow the possibility that absolute ungrammaticality is not located in the H-Eval component of grammar, but in a component that precedes (Gen) or follows (interpretation) optimization. If, however, H-Eval is to be held responsible for the ungrammaticality of (82), there must be a competing candidate with a better constraint profile that blocks it. A priori, this might be a candidate that employs a resumptive pronoun strategy, which is only legitimate in this context as a last resort. If this were so, the ineffability problem would be spurious in the case at hand. However, (85) shows that the resumptive pronoun strategy is not an option in German (a constraint like RES must outrank ADJ-ISL and other locality constraints in German):

(85)   *Was$_1$ ist Fritz eingeschlafen [$_{CP}$ nachdem er es$_1$ gelesen hat] ?
           what is Fritz fallen asleep   after   he it   read   has

What, then, could the optimal candidate blocking (82) look like? Following Prince & Smolensky (1993), Ackema & Neeleman (1998) propose that the empty candidate Ø (the "null parse") is part of every candidate set. This candidate violates the constraint in (86), which is typically ranked high.[40]

(86)   *Ø ("Avoid Null Parse"):
           Ø is prohibited.

Constraints that are ranked higher than *Ø in effect become inviolable (given that there is no constraint except *Ø that Ø can violate). In this sense, *Ø introduces a dividing line into rankings. Thus, if both ADJ-ISL and the constraint that triggers *wh*-movement (e.g., OP-SPEC) outrank *Ø, adjunct islands become inviolable. This is shown in table T$_{12}$.

T$_{12}$: Adjunct islands and the null parse

Candidates	ADJ-ISL	OP-SPEC	*Ø
C$_1$: was$_1$ ... [$_{CP}$ nachdem er t$_1$ V]	*!		
C$_2$: — ... [$_{CP}$ nachdem er was$_1$ V]		*!	
☞ C$_3$: Ø			*

A final possibility to be discussed here is the *neutralization* approach to absolute ungrammaticality in syntax. Such an approach has been adopted by Legendre, Smolensky & Wilson (1998), Schmid (1998), Baković & Keer (1999), and Wilson (1999), among others. For the present case, a neutralization analysis might posit that the optimal candidate blocking (82) is (87).

(87)   Fritz ist eingeschlafen [$_{CP}$ nachdem er was$_1$ gelesen hat]
        Fritz is fallen asleep   after   he something read   has

The crucial difference from (84) is that *was$_1$* is turned into an indefinite pronoun, and the matrix C$_{[+wh]}$ is turned into a C$_{[-wh]}$. Thus, there is a feature change from [+wh] in (82) to [–wh] in (87), and the sentence is interpreted as declarative, rather than as a question.[41] If (87) is to block (82) as suboptimal, this presupposes that candidates that differ in their *wh*-feature specification can compete. But then, the problem arises that we would also wrongly expect one of the sentences in (88) to block the other.

(88) a. Was$_1$ hat er t$_1$ gelesen ?
         what has he   read
     b. Er hat was$_1$ gelesen
         he has something read

The neutralization approach solves this problem as follows. The [±wh]-specification is unambiguously specified in the input; an input with a [+wh] specification on some item and a minimally different input with a [–wh] specification count as different, and define different candidate sets. The important assumption is that there is a faithfulness constraint that demands preservation of the [±wh] feature specification in the output:

(89)   FAITH[WH]:
        The output value of [±wh] is the same as the input value.

Suppose now that ADJ-ISL and OP-SPEC are ranked higher than FAITH[WH]. Then, (87) will have a better constraint profile than (82) both in the competition that has a [–wh] specification in the input, and in the competition that has a [+wh] specification in the input. Thus, there is a "neutralization" of different input specifications in the output. This is shown in tables T$_{13}$ and T$_{14}$.[42]

T_{13}: Adjunct islands and neutralization; [–w] in the input

Candidates	ADJ-ISL	OP-SPEC	FAITH[WH]
C_1: was$_{1[+w]}$ ... [$_{CP}$ nachdem er t$_1$ V]	*!		*
C_2: — ... [$_{CP}$ nachdem er was$_{1[+w]}$ V]		*!	*
☞ C_3: — ... [$_{CP}$ nachdem er was$_{1[-w]}$ V]			

T_{14}: Adjunct islands and neutralization; [+w] in the input

Candidates	ADJ-ISL	OP-SPEC	FAITH[WH]
C_1: was$_{1[+w]}$ ... [$_{CP}$ nachdem er t$_1$ V]	*!		
C_2: — ... [$_{CP}$ nachdem er was$_{1[+w]}$ V]		*!	
☞ C_3: — ... [$_{CP}$ nachdem er was$_{1[-w]}$ V]			*

In transparent contexts, where movement may occur without a violation of a high-ranked locality constraint like ADJ-ISL (cf. (88)), FAITH[WH] violations become fatal, and the candidate that maintains the [±wh] specification of the input emerges as optimal.[43]

Of the four approaches to absolute ungrammaticality discussed here (Gen, interpretation, null parse, neutralization), the neutralization approach is arguably the most elegant one. Still, it is not without problems. One conspicuous peculiarity is that neutralization creates massive derivational ambiguity. A well-formed sentence like (87) can have different "histories," being an optimal candidate in two candidate sets with different inputs. This vacuous ambiguity may be considered problematic from the point of view of language acquisition and parsing; and it can only be avoided by additional meta-optimization procedures that compare the competitions in T_{13} and T_{14}; cf. the notion of *input optimization* in Prince & Smolensky (1993) (called *lexicon optimization* in phonology).

5.3.3 Residual Issues

As remarked above, this does not exhaust the list of open issues that are currently under debate in optimality-theoretic syntax. We end this section by briefly mentioning a few others.

Optionality

In the best of all possible worlds, one would not expect optionality to arise in a theory that selects the best candidate. The solutions that have been proposed in view of this situation center around concepts like (i) *true optionality*, according to which more than one candidate can be optimal due to an identical constraint profile (recall the above discussion of complementizer-trace effects); (ii) *constraint ties*, which come in various versions (global and local, ordered, conjunctive, and disjunctive) and all somehow incorporate the idea that two (or more) constraints are equally important; (iii) *pseudo-optionality*, which rests on the idea that the observed optionality is only apparent, and reducible to different optimization procedures in different candidate sets; and (iv) *neutralization* again, essentially an elaborate version of (iii). It turns out that none of these solutions is completely unproblematic. See Müller (2000:chapter 5) for a critical overview.

Degrees of Grammaticality

According to the definition of optimality in (70), an optimal candidate is grammatical, and a suboptimal candidate is invariably ungrammatical, no matter what the relative quality of its constraint profile is in comparison with other suboptimal candidates. Without further assumptions, this makes it impossible to account for degrees of grammaticality (or acceptability) in a syntax-internal way, in contrast to what is the case in government and binding theory (albeit only by stipulation; cf., e.g., the traditional distinction between "mild" Subjacency violations and "strong" ECP violations).

Cumulativity

A related property of optimality-theoretic syntax is that, in its standard form, it does not capture cumulative effects; in government and binding theory, cumulativity manifests itself in the assumption that a sentence gets "more ungrammatical," the more constraints it violates. The reason for optimality theory's failure to integrate cumulativity is that many violations of a lower-ranked constraint cannot outweigh a single violation of a higher-ranked constraint. However, as we have seen, this consequence does not hold if we adopt the mechanism of local constraint conjunction. Whether this is a positive or negative result remains to be seen.

Parameterization

Work in government and binding theory and the minimalist program has focussed on morphological properties of lexical items as factors that determine parametrization. Such a view can in principle be reconciled with optimality-theoretic syntax without too much ado (one and the same syntactic constraint ranking may yield different optimal candidates if the morphological properties of these candidates differ from language to language, and there are constraints that refer to these morphological properties). However, in practice, work in optimality-theoretic syntax has often sought to account for syntactic parameterization exclusively in terms of syntactic reranking, and either deny a relation to morphology, or view morphological properties not as the basis, but as a reflex of syntactic parameterization. Again, this issue is far from being settled; for opposing views, see, e.g., Grimshaw & Samek-Lodovici (1998) and Legendre, Smolensky & Wilson (1998) on the one hand, and Vikner (2000) on the other.

   Another recurring question in the optimality-theoretic approach to parameterization is whether every reranking of constraints that is logically possible is also linguistically plausible (i.e., results in a potential grammar). The hypothesis that it is is known as *factorial typology*, and is the focus of much recent work.

Multiple Optimization

Following Prince & Smolensky (1993), it is standardly assumed that there is exactly one optimization procedure in syntax; the candidates are evaluated only once. An alternative that is considered in Prince & Smolensky (1993) is that optimization procedures can affect candidates more than once. Recently, this idea has been pursued in various ways in optimality-theoretic syntax. Several proposals rely on the distinction between interpretive optimization and expressive optimization: Interpretive optimization may precede expressive optimization (see Wilson 1999), expressive optimization may precede interpretive optimization (see Hendriks & de Hoop 1999), or the two procedures may influence each other (see Blutner 2000 and Jäger & Blutner 2000). Heck (1998:this volume) argues that the government and binding model can be transferred into optimality-theoretic syntax by assuming that optimization applies three times: First, D-structures are subject to optimization; second, the optimal D-structure output serves as the input to S-structure optimization; finally, the optimal S-structure output serves as the input to LF optimization.

Fanselow, Kliegl & Schlesewsky (1999) develop an optimality-theoretic approach to parsing that is based on the idea that parsing can be viewed as an iteration of optimization procedures that stop when the final word of a sentence has been taken in. Finally, Heck & Müller (2000) adopt a minimalist syntax in which each cyclic node (XP) created in the derivation is subject to optimization; only the optimal XP is submitted to the next step of the derivation, and so on, until the optimal root node is determined. Thus, in this system, optimization is not just multiple; it is local in the sense that each optimization procedure affects only a small unit.

   None of these cases of multiple optimization can be viewed as a notational variant of standard, single optimization. It remains to be seen to what extent multiple optimization is a viable alternative.

6   The Contributions to This Volume

Most of the papers in this volume originate from a workshop at the 21st Annual Conference of the DGfS (German Linguistic Society), which took place at the university of Constance in February, 1999. The contributions have in common that they discuss pieces of empirical evidence for which a competition-based approach has some initial plausibility. They are all primarily concerned with optimality theory, and they take up a number of the open issues that were just mentioned.

   Büring's paper is a study of free word order in German, a domain that has been tackled in terms of violable and ranked constraints in pre-optimality work going back to the 70's and 80's. Like Choi (1999), Büring's approach presents an optimality-theoretic analysis that rests on Lenerz's (1977) seminal work. Central theoretical notions that play a role include optionality, degrees of grammaticality, and, in particular, the prosody/syntax interface.

   Fanselow & Ćavar adopt the copy theory of movement and assume that overt and covert movement both apply before spell-out. The crucial difference relates to the question of which members of a copy chain are pronounced, and which are deleted. To give a comprehensive answer to this question, the authors discuss evidence from a variety of languages that includes long-distance and partial *wh*-movement, the *wh*-copy construction, the NP split construction, and instances of head movement. They develop an optimality-theoretic approach that reconciles features of the analyses in Pesetsky (1998) and Grimshaw (1997), and that relies on a system of multiple (local) optimization which integrates Chomsky's (1998) concept of a *phase*.

As in the case of word order, it has often been argued in pre-optimality-theoretic analyses of relative quantifier scope that the notions of violability and ranking (or weight) of interacting factors play an important role. **Fischer** sets out to transfer some main results of one such study (viz., Pafel's 1998 approach to quantifier scope in German) into optimality theory. In view of the fact that this approach also employs the notions of optionality (in the guise of scope ambiguity) and cumulativity, Fischer develops an analysis that rests on constraint ties and local constraint conjunction.

Heck is also concerned with quantifier scope in German. Based on the observation that scope relations at LF are highly dependent on word order at S-structure, which is in turn strongly influenced by the variable order of arguments at D-structure, he argues for a new system of multiple optimization (called "cyclic optimization"). This system takes the government and binding organization of grammar as a starting point and postulates three optimization procedures: at D-structure, at S-structure, and at LF.

In his experimental study of English gapping constructions, **Keller** observes the influence of various interacting constraints. On the basis of this evidence, he argues for an approach that incorporates features of optimality-theoretic syntax but also provides room for (a) cumulativity of constraint violation; (b) gradient acceptability of candidates (related to the number and quality of constraint violations); and (c) a distinction between "hard" and "soft" constraints (e.g., a hard clause-mateness constraint, and soft subject/predicate and minimal distance constraints), the latter being violable and subject to choice of context.

Like Büring, **Lenerz** addresses free word order structures in German. This paper complements Büring's, since Lenerz argues that the empirical evidence does not in fact support a competition-based approach. Going through all the main pieces of word order evidence that have been analyzed in terms of competition, Lenerz shows that an analysis that focusses on the variable semantic and pragmatic contributions of definite and indefinite NPs in different positions in the German middle field can yield empirically adequate results without any recourse to the notion of competition; the particular approach that he develops relies on choice functions and the partitioning of clauses into domains with background-determined reference and with immediate sentence constituent reference.

Schmid's paper is a close investigation of different ways to handle optionality in optimality-theoretic syntax. After reviewing the options that exist in the literature, Schmid focusses on a comparison of a specific (global) notion of constraint tie and the concept of neutralization. The cases of optionality that

serve as the empirical basis are (a) complementizer drop in English, (b) *wh*-movement in French root clauses, and (c) the German "Ersatzinfinitiv" (IPP) construction. For each of these phenomena, a global tie analysis is compared with a neutralization analysis; general strategies are suggested that permit a transfer from one type of approach to the other; and a conclusion is drawn that ultimately favors the neutralization solution.

The focus of **Vikner**'s paper is the conflict that arises between two well-motivated constraints in Icelandic: First, the relative scope of quantified items must correspond to their surface order; second, NPs can undergo object shift in front of an adverbial only if the main verb has undergone movement. Interestingly, it seems as though relative scope does not have to correspond to surface order in exactly those contexts in which object shift is blocked. Vikner shows that this supports an optimality-theoretic analysis in which the first constraint is ranked below the second one, and is thus violable in the case of conflict. Finally, the analysis is extended to German.

Vogel takes as a starting point the observation that free relative constructions by their very nature strongly suggest constraint violability and constraint ranking: They are incompatible with the standard assumption that there is a one-to-one correspondence between Case assigners and items that are assigned Case. Moreover, Case conflicts can show up in free relatives which are often resolvable by ranking (but may also result in absolute ungrammaticality). On this basis, Vogel develops an optimality-theoretic analysis of free relative constructions in German, and he investigates the typological implications that result from reranking the proposed constraints; among other things, the analysis sheds new light on the concepts of factorial typology and neutralization.

Finally, **Wanner** observes that there are conflicts between linking rules which become manifest in the domain of psych verbs in English. For instance, the CONTROL-RULE favors experiencers as external arguments, whereas the CAUSER-RULE prefers causers as external arguments; in an optimality-theoretic approach, the conflict can be resolved by ranking the latter rule above the former, and this is what explains the difference between *Mary frightens John* (where the theme is a causer) and *John fears Mary* (where the theme is not a causer). An interesting theoretical aspect of this analysis is that the competing candidates are not sentences, but argument structures.

We believe that the papers collected in this volume give a fair indication of both the potential and the limitations of optimality-theoretic syntax, and of competition-based syntax in general. To us, they strongly suggest that it is

fruitful to further explore the concept of syntactic competition, even though an eventual success of this enterprise cannot be taken for granted at this point.

Acknowledgments

Notes

1. Our use of the term *global* follows its original interpretation in Lakoff (1971) throughout this introduction. Sometimes, *global* is understood in a rather different sense in the literature (including Chomsky 1995 and Collins 1997), as a synonym for *translocal* or *transderivational* (see below). As we will see, in this second interpretation, a global constraint can in fact not be checked by exclusively looking at a given syntactic object S_i.
2. The resumptive pronoun strategy is by itself marginal in English and is chosen here mainly for expository reasons; see Chomsky (1981:173) for a discussion of the case at hand. However, resumptive pronouns as a last resort in cases where movement is blocked are widely attested in other languages. See Shlonsky (1992), Pesetsky (1998), and the references cited there.
3. Note, however, that Chomsky (1982:63f.) envisages an account in terms of the Avoid Pronoun Principle, which, as we will see, is an exception insofar as it is in fact a non-local constraint in government and binding theory.
4. As Manzini shows, the Control Rule is actually a theorem that can be derived from more primitive assumptions. This need not concern us here.
5. The Avoid Pronoun Principle has been applied to pro-drop phenomena in languages like Italian by Haegeman (1994:217). The idea here is that the availability of the empty pronominal pro in the subject position of finite clauses tends to make the use of an overt pronoun impossible; on this view, overt subject pronouns can only show up in pro-drop languages if they fulfill a function that pro cannot fulfill (like, e.g., focus interpretation). Also recall Chomsky's (1982) analysis of resumptive pronouns that was mentioned above.
6. Also see Reinhart (1983) on a version of binding theory that relies on pragmatic constraints of this type.

7. The notion of a numeration is first introduced in Chomsky (1993), so this is strictly speaking an anachronism.

8. This view is later abandoned in the minimalist program. Thus, Chomsky (1998:6) speculates that "language design might be optimal ... approaching a 'perfect solution' to minimal design specifications."

9. We can assume that a position is "appropriate" for insertion of intermediate traces if the resulting structure does not violate local constraints – e.g., those on improper movement (an A-bar trace must not end up within an A-chain, see May 1979; an adjoined trace must not end up within a chain headed by an antecedent in a specifier position (and vice versa), see Müller & Sternefeld 1996).

10. At least, this holds as long as we are not prepared to assume that embedded V/2 in German is derived by deletion of a complementizer *daß* that is present in the numeration.

11. All XPs which are not L-marked are barriers. XPs which are not in complement position are therefore always barriers.

12. Note that the concept of Form Chain cannot undermine this reasoning because the two instances of chain formation applying to the *wh*-phrase in D_3 are not adjacent, but interrupted by another operation – that of NP raising to subject position.

13. See Chomsky (1993, 1995) for discussion of the various options that arise here.

14. Also compare Kitahara's (1997) reconstruction of Procrastinate effects in terms of Fewest Steps; see section 3.3 on Procrastinate.

15. Note in passing that Chomsky's (1991) way out in terms of 'stylistic movement' that can be chosen in the case of optional overt movement is not viable here for obvious reasons, LF movement never being 'stylistic.'

16. The derivation originally envisaged by Collins (1994) is actually even more complex since it involves two additional VP-adjunction operations. The derivation in (46) is sufficient for our present purposes, though.

17. One might think that Fewest Steps would also suffice to block D_1 in favor of D_2. However, assuming that the two movement operations in D_1 can be reanalyzed as a single instance of Form Chain, this is not the case. That said, it is worth noting that Shortest Paths would indeed suffice to account for the ban on V-in situ in French that was explained by invoking Fewest Steps in Chomsky (1991). The derivation in (17) also instantiates yo-yo movement; the only essential difference from the derivation in (46) is that yo-yo movement is interrupted by spell-out in the former case, but not in the latter.

18. Two additional assumptions must be clarified. First, in line with what is probably the majority of literature on the topic, Nakamura (1998) postulates that *wh*-constructions in Tagalog do not actually involve movement of the *wh*-phrase, but rather movement of an empty operator in a relative clause-like construction. That is, English questions like "What did Juan buy?" are rendered as "What is it

Op_1 that Juan bought t_1?" For expository purposes, we will ignore this complication in what follows, but the correct structure is still reflected in the translation. Second, note that actual positions of items that are overtly visible do not always reflect the position that is theoretically relevant in Nakamura's (1998) analysis. In particular, he assumes that the structural subject position SpecT is left-peripheral, and in many cases can only be filled at LF. Still, subject NPs behave in every respect as if they occupied the SpecT position overtly. This covert subject raising with overt effects is indicated here by italicizing the relevant subject NP; thus italicization is meant to imply that the italicized NP is pronounced in the position of its trace. Note that these complications do not arise in a language like Toba Batak, which otherwise exhibits the same general effect; see Schachter (1984) and Sternefeld (1995).

19. Note, however, that a residue of the Procrastinate condition still shows up in Chomsky's (1998:14) translocal principle that prefers Agree over Move. Also see the next section.

20. It is worth noting that the notion of optimality has systematically been used in minimalist syntax, apparently without recourse to optimality theory as developed by Prince & Smolensky (1993), and at a time when optimality-theoretic syntax papers did not yet exist. See, e.g., Chomsky (1993:4), and, for explicit uses of the notion, Collins (1994:46), Kitahara (1997:18), and Frampton & Gutman (1999:5).

21. See Müller (2000:chapter 4) for a slightly more realistic (albeit still simplified) example.

22. See Fanselow (1997), who argues that was_2 can be scrambled to a position in front of wer_1 before *wh*-movement takes place in (61-b). However, to avoid a blocking of (61-a) by (61-b) in a system with translocal constraints (which Fanselow does not assume), it would then also have to be ensured that the two derivations do not compete; this could be achieved by assuming that the presence vs. absence of the trigger for *was*-scrambling creates two different reference sets. An alternative would be to assume that whereas translocal constraints cannot be parameterized, the definition of reference set can be. Without recourse to intermediate scrambling, reference sets might then be defined in German in such a way that (61-a) and (61-b) do not compete, whereas they could be defined differently in English, so that the English counterparts of these derivations do compete. See Sternefeld (1997) for an extensive discussion of this option.

23. Recall, however, that Chomsky retains some translocal constraints even in more recent work, though often hesitantly and with a sense that if truly necessary, translocality would qualify as an "imperfection" of language. Thus, directly after suggesting the Shortest Paths account of the ban on the acyclic derivation of freezing effects with NP raising cited above, Chomsky (1995, 328) comes close to revoking it by stating: "– though the issue is nontrivial, in part because we are

invoking here a 'global' [i.e., translocal] notion of economy of the sort we have sought to avoid."

24. Also see Hornstein (2000), who gives the same kind of account in a minimalist setting.

25. The standard way out of the problems created by optionality chosen by proponents of blocking syntax is to find subtle semantic differences between the relevant sentences – in other words, to deny true optionality.

26. The former strategy is discussed in Heck (1998:this volume); the latter strategy is pursued in Müller (1997).

27. We hasten to add that all the case studies in this section are simplified versions of the actual analyses proposed in the literature. In the present context, we are mainly interested in the logic of the argument, not in the specific (or maximally elegant) formulation of the constraints. Accordingly, we leave open the questions of defining candidates and candidate sets where they do not seem to be important for our present purposes. Note also that the simplification is particularly radical in Wilson's (1999) case. Based on evidence from binding theory, Wilson argues for an elaborate model of multiple optimization in syntax (see section 5.3.3 below); he is concerned with many more data and, eventually, typological universals that the naive analysis presented here cannot possibly account for.

28. The ranking of T-Lex-Gov and Stay could also be reversed; this ranking is not determined by the cases we are interested in here.

29. To avoid the issue of *do*-support in root clauses, which is orthogonal to the issue of complementizer-trace effects in embedded clauses, we have chosen examples here in which the $\text{SpecC}_{[+wh]}$ target position is in an embedded clause.

30. Note that a violation of T-Lex-Gov will automatically imply a violation of the more general Stay constraint. Hence, given that there is no other constraint on which C_1 and C_2 differ, it follows that C_1's constraint profile is better than that of C_2 under ranking; i.e., C_1 *harmonically bounds* C_2

31. The relevant constraints are Bar^3 (see below) and Fill in Legendre, Smolensky & Wilson (1998), and Island-Cond and Silent-Trace in Pesetsky (1998). Cnpc should be viewed as a placeholder for one or more general conditions that yield the described effects; Res is arguably part of a more general system of constraints on pronouns. Also see Hornstein (2000) resumptive pronouns and islands, including CNPC, in English.

32. Operator movement in relative clauses in English can be achieved by (something along the lines of) Grimshaw's (1997) ranking Op-Spec ≫ Stay (plus Stay ≫ Res); this option is chosen by Legendre, Smolensky & Wilson (1998). In contrast, Pesetsky (1997) does not assume the movement operation in relative clauses to be subject to optimization; in his view, Gen does not generate the in-situ version in English in the first place.

33. It seems that in order to achieve compatibility of this account of resumptive pronouns with the account of the lack of complementizer-trace effects with adjuncts

sketched in the preceding section, the ranking RES ≫ T-LEX-GOV would have to be assumed. That said, one will probably have to assume independent, high-ranked constraints that block resumptive pronouns in adjunct chains, anyway.

34. Is *PRON confined to personal (and possessive) pronouns, or does it also cover anaphoric pronouns? Under the first option, REF-ECON and *PRON might be the same constraint. Under the second option, we would in fact face what is known as a *subhierarchy* of constraints: A general constraint *PRON prohibits all kinds of pronouns, a more specific constraint *PERS-PRON (= REF-ECON) prohibits only personal pronouns, and an even more specific constraint *RES-PERS-PRON (= RES) prohibits only personal pronouns used as resumptives.

35. Ackema & Neeleman call this constraint STAY, but this may be somewhat unfortunate, given that MIN-CHAIN differs substantially from Grimshaw's (1997) STAY, in the same way that Shortest Paths differs from Fewest Steps.

36. Local constraint conjunction makes it possible to reintroduce the concept of cumulativity into optimality-theoretic syntax: Multiple violations of a given constraint Con_1 may not directly outweigh a single violation of a higher-ranked constraint Con_2, but can do so indirectly by triggering a violation of an even higher-ranked constraint $Con_1{}^i$.

37. An interesting question is whether a translation of translocal constraints into local constraints is actually needed in optimality theory; in other words: Could not some of the violable and ranked constraints in the H-Eval part be translocal themselves, just like the basic optimality principle is? For instance, one could envisage a translocal SHORTEST PATHS that fulfills the same task as MIN-CHAIN or BAR^i: SHORTEST PATHS selects the candidate C_i with the shortest movement paths in a given candidate set, and this can be signalled by stipulating that all candidates except C_i are assigned a star * under this constraint. Such an approach may raise additional complexity issues, and it has – to the best of our knowledge – not yet been proposed in optimality-theoretic syntax. Still, it seems to us to be viable in principle. Indeed, a translocal constraint of this type has been proposed for phonology in Prince & Smolensky (1993) (H-NUC, which, however, is eventually replaced there by a subhierarchy of local constraints that are derived by a process of *harmonic alignment*).

38. In both cases, only those aspects of the input are considered that matter for the faithfulness constraints under consideration.

39. Note that the distinction between the actual scope position for a *wh*-item (here designated by [+wh]) and the "intended" scope position for a *wh*-item (here designated by Σ) is fundamental in the analysis by Legendre, Smolensky & Wilson (1998), and not an artefact of the input-independent approach.

40. Assuming the concept of input, this constraint amounts to the statement that the input must not be left completely unrealized.

41. Here we exploit the fact that *was* is ambiguous between a *wh*-reading and an indefinite reading in (colloquial) German. This does not hold for other *wh*-phrases

like *welches Buch* ('which book'), which, however, also cannot be extracted from adjunct islands. For these cases, the neutralization approach would have to be complicated in such a way that the candidate with the [–wh] NP (perhaps *ein Buch* ('a book')) deviated from the one that must be blocked not just in feature specification, but also in morphological shape. Such complications do not affect the general argument, though.

42. A side remark: The candidates in (79-b) and (79-d) that were discussed in the previous section also signal input neutralization; these candidates are also optimal in candidate sets where they do not violate the respective faithfulness constraint.

43. This account rests on the concept of input. Is it possible to maintain the analysis without reference to this notion? It is, but the task is slightly more difficult here than in the cases that were discussed in the last section. We have to ensure that an output candidate like (87), with a $C_{[-wh]}$ and a $was_{[-wh]}$, has abstract [+wh] or [–wh] markers that encode the postulated input difference, and that can be referred to by an appropriately revised FAITH[WH] constraint.

References

Ackema, Peter — Ad Neeleman
　　1998　Optimal questions. *Natural Language and Linguistic Theory* 16: 443-490.

Archangeli, Diana — D. Terence Langendoen
　　1996　Afterword. In: Diana Archangeli & D. Terence Langendoen (eds.), *Optimality Theory: An Overview*, 200-215. Oxford: Blackwell.

Aronoff, Mark
　　1976　*Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Baker, Carl L.
　　1970　Notes on the description of English questions: The role of an abstract question morpheme. *Foundations of Language* 6: 197-219.

Baković, Eric — Ed Keer
　　1999　Optionality and ineffability. Ms., Harvard University & UMass., Amherst. To appear in: Géraldine Legendre, Jane Grimshaw & Sten Vikner (eds.), *Optimality-Theoretic Syntax*, Cambridge, MA: MIT Press.

Blutner, Reinhard.
　　2000　Some aspects of optimality in natural language interpretation. Ms., Humboldt-Universität Berlin.

Burzio, Luigi
　　1991　The morphological basis of anaphora. *Journal of Linguistics* 27: 81-105.

Choi, Hye-Won
   1999   *Optimizing Structure in Context: Scrambling and Information Structure.* Stanford: CSLI Publications.

Chomsky, Noam
   1973   Conditions on transformations. In: Stephen Anderson & Paul Kiparsky (eds.), *A Festschrift for Morris Halle*, 232-286. New York: Academic Press.

Chomsky, Noam
   1981   *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam
   1982   *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, MA: MIT Press.

Chomsky, Noam
   1986a   *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam
   1986b   *Knowledge of Language*. New York: Praeger.

Chomsky, Noam
   1991   Some notes on economy of derivation and representation. In: Robert Freidin (ed.), *Principles and Parameters in Comparative Grammar*, 417-454. Cambridge, MA: MIT Press.

Chomsky, Noam
   1993   A minimalist program for linguistic theory. In: Kenneth Hale & Samuel Jay Keyser (eds.), *The View from Building 20*, 1-52. Cambridge, MA: MIT Press.

Chomsky, Noam
   1995   Categories and transformations. (Chapter 4). In: *The Minimalist Program*, 219-394. Cambridge, MA: MIT Press.

Chomsky, Noam
   1998   Minimalist inquiries. Ms., MIT, Cambridge, MA

Chomsky, Noam — Howard Lasnik
   1993   Principles and parameters theory. In: Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax, vol. I*, 506-569. Berlin: de Gruyter.

Cole, Peter
   1982   Subjacency and successive cyclicity: Evidence from Ancash Quechua. *Journal of Linguistic Research* 2: 35-58.

Collins, Chris
   1994   Economy of derivation and the generalized proper binding condition. *Linguistic Inquiry* 25: 45-61.

Collins, Chris
    1997   *Local Economy*. Cambridge, MA: MIT Press.

Déprez, Viviane
    1991   Economy and the *that*-t effect. In *Proceedings of the Western Conference on Linguistics* 4: 74-87.

DiSciullo, Anna-Maria — Edwin Williams
    1987   *On the Definition of Word*. Cambridge, MA: MIT Press.

Epstein, Samuel David
    1992   Derivational constraints on A′–chain formation. *Linguistic Inquiry* 23: 235-259.

Fanselow, Gisbert
    1989   Konkurrenzphänomene in der Syntax. *Linguistische Berichte* 123: 385-414.

Fanselow, Gisbert
    1991   Minimale Syntax. Habilitation thesis, Universität Passau.

Fanselow, Gisbert
    1997   The proper interpretation of the minimal link condition. Ms., Universität Potsdam.

Fanselow, Gisbert — Reinhold Kliegl — Matthias Schlesewsky
    1999   Optimal parsing. Ms., Universität Potsdam.

Fox, Danny
    1995   Economy and scope. *Natural Language Semantics* 3:283-341.

Frampton, John — Sam Gutman
    1999   Cyclic computation. *Syntax* 2: 1-27.

Grimshaw, Jane
    1994   Heads and optimality. Handout, Universität Stuttgart.

Grimshaw, Jane
    1997   Projection, heads, and optimality. *Linguistic Inquiry* 28: 373-422.

Grimshaw, Jane — Vieri Samek-Lodovici
    1998   Optimal subjects and subject universals. In: Pilar Barbosa et al. (eds.), *Is the Best Good Enough?*, 193-219. Cambridge, MA: MIT Press & MITWPL.

Haegeman, Liliane
    1994   *Introduction to Government and Binding Theory*. Oxford: Blackwell.

Haider, Hubert
    1983   Connectedness effects in German. *Groninger Arbeiten zur Germanistischen Linguistik* 23: 82-119.

Heck, Fabian
  1998   Relativer Quantorenskopus im Deutschen – Optimalitätstheorie und die Syntax der Logischen Form. M.A. thesis, Universität Tübingen.

Heck, Fabian — Gereon Müller
  2000   Repair-driven movement and the local optimization of derivations. Ms., Universität Stuttgart & IDS Mannheim. Short version in: *Glow Newsletter* 44: 26-27.

Hendriks, Petra — Helen de Hoop
  1999   Optimality theoretic semantics. Ms., University of Groningen. (Cognitive Science and Engineering Prepublications 98-3.)

Hornstein, Norbert
  2000   Is the binding theory necessary? Ms., University of Maryland.

Jäger, Gerhard — Reinhard Blutner
  2000   Against lexical decomposition in syntax. Ms., ZAS & Humboldt-Universität Berlin.

Kiparsky, Paul
  1982   From cyclic phonology to lexical phonology. In: Harry van der Hulst & Neil Smith (eds.), *The Structure of Phonological Representations, vol 1*, 131-175. Dordrecht: Foris.

Kitahara, Hisatsugu
  1993   Deducing 'superiority' effects from the shortest chain requirement. *Harvard Working Papers in Linguistics* 3: 109-119.

Kitahara, Hisatsugu
  1997   *Elementary Operations and Optimal Derivations*. Cambridge, MA: MIT Press.

Koster, Jan
  1987   *Domains and Dynasties*. Dordrecht: Foris.

Lakoff, George
  1971   On generative semantics. In: Danny Steinberg & Leon Jakobovits (eds.), *Semantics*, 232-296. Cambridge: Cambridge University Press.

Lasnik, Howard — Mamoru Saito
  1992   *Move α*. Cambridge, MA: MIT Press.

Legendre, Géraldine — Paul Smolensky — Colin Wilson
  1998   When is less more? Faithfulness and minimal links in wh-chains. In: Pilar Barbosa et al. (eds.), *Is the Best Good Enough?*, 249-289. Cambridge, MA: MIT Press & MITWPL.

Lenerz, Jürgen
  1977   *Zur Abfolge nominaler Satzglieder im Deutschen*. Tübingen: Stauffenburg.

Manzini, Rita
   1983   On control and control theory. *Linguistic Inquiry* 14: 421-446.

Marantz, Alec
   1995   The minimalist program. In: Gert Webelhuth (ed.) , *Government and Binding Theory and the Minimalist Program*, 351-382. Oxford: Blackwell.

May, Robert
   1979   Must COMP-to-COMP movement be stipulated? *Linguistic Inquiry* 10: 719-725.

McCarthy, John — Alan Prince
   1995   Faithfulness and reduplicative identity. In: Jill Beckman, Laura Walsh-Dickie & Suzanne Urbanczyk (eds.) , *Papers in Optimality Theory*, 249-384. Amherst, MA: UMass Occasional Papers in Linguistics 18.

Müller, Gereon
   1997   Partial wh-movement and optimality theory. *The Linguistic Review* 14: 249-306.

Müller, Gereon
   2000   *Elemente der optimalitätstheoretischen Syntax*. Tübingen: Stauffenburg.

Müller, Gereon — Wolfgang Sternefeld
   1996   A-bar chain formation and economy of derivation. *Linguistic Inquiry* 27: 480-511.

Nakamura, Masanori
   1998   Reference set, minimal link condition, and parameterization. In: Pilar Barbosa et al. (eds.), *Is the Best Good Enough?*, 291-313. Cambridge, MA: MIT Press & MITWPL.

Pafel, Jürgen
   1998   Skopus und logische Struktur — Studien zum Quantorenskopus im Deutschen. Habilitationsschrift, Universität Tübingen.

Pesetsky, David
   1997   Optimality theory and syntax: Movement and pronunciation. In: Diana Archangeli & D. Terence Langendoen (eds.), *Optimality Theory. An Overview*, 134-170. Oxford: Blackwell.

Pesetsky, David
   1998   Some optimality principles of sentence pronunciation. In: Pilar Barbosa et al. (eds.), *Is the Best Good Enough?*, 337-383. Cambridge, MA: MIT Press & MITWPL.

Pollock, Jean-Yves
   1989   Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry* 30: 365-424.

Prince, Alan — Paul Smolensky
   1993   *Optimality Theory: Constraint Interaction in Generative Grammar*. Ms., Rutgers University. To appear: Cambridge, MA: MIT Press.

Reinhart, Tanya
   1983   *Anaphora and Semantic Interpretation*. London: Croom Helm.

Richards, Norvin
   1997   Competition and disjoint reference. *Linguistic Inquiry* 28: 178-187.

Schachter, Paul
   1984   *Studies in the Structure of Toba Batak*. UCLA Occasional Papers in Linguistics 5.

Schmid, Tanja
   1998   West germanic "Infinitivus Pro Participio" (IPP) constructions in optimality theory. In: Tina Cambier-Langeveld, Anikó Lipták, Michael Redford & Erik Jan van der Torre (ed.) , *Proceedings of Console VII*, 229-244. Leiden: SOLE.

Shlonsky, Ur
   1992   Resumptive pronouns as a last resort. *Linguistic Inquiry* 23: 443-468.

Speas, Margaret
   1995   Generalized control and null objects in optimality theory. In: Jill Beckman, Laura Walsh-Dickie & Suzanne Urbanczyk (eds.) , *Papers in Optimality Theory*, 637-653. Amherst, MA: UMass Occasional Papers in Linguistics 18.

Sternefeld, Wolfgang
   1991   Chain formation, reanalysis, and the economy of levels. In: Hubert Haider & Klaus Netter (eds.), *Representation and Derivation in the Theory of Grammar*, 71-137. Dordrecht: Kluwer.

Sternefeld, Wolfgang
   1995   Voice phrases and their specifiers. *FAS Papers in Linguistics* 3: 48-85.

Sternefeld, Wolfgang
   1997   Comparing reference sets. In: Chris Wilder, Hans-Martin Gärtner & Manfred Bierwisch (eds.) , *Economy in Linguistic Theory*, 81-114. Berlin: Akademieverlag.

Vikner, Sten.
   2000   Checking strong verbal inflection in optimality theory. Ms., Universität Stuttgart.

Williams, Edwin
   1986   A reassignment of the functions of LF. *Linguistic Inquiry* 17: 265-299.

Williams, Edwin
   1997   Blocking and anaphora. *Linguistic Inquiry* 28: 577-628.

Wilson, Colin
   1999   Bidirectional optimization and the theory of anaphora. Ms., Johns Hop-
           kins University. To appear in: Géraldine Legendre, Jane Grimshaw &
           Sten Vikner (eds.) *Optimality Theoretic Syntax*, Cambridge, MA: MIT
           Press.