Minimal links, remnant movement, and
(Non-)derivational grammar

John Hale and Géraldine Legendre

1. Introduction

The Minimal Link Condition of Chomsky (1995, 311) exemplifies a particular
perspective on the role of optimization in grammar.

(1) Minimal Link Condition (MLC)

K attracts α only if there is no β, β closer to K than α, such that K
attracts β

On this view, optimization plays a role in selecting operations that apply to
intermediate stages of grammatical derivations. The role is inherently local in
character, suggesting a “derivational” view of grammar as a kind of machine
taking linguistic rules as instructions. Chomsky has adopted this view as part
of the Minimalist Program (MP, 1999, 12):

One might construe L as a step-by-step procedure for constructing EXPs,
suggesting that this is how things work is a real property of the brain, not
temporally but as part of its structural design. Assumptions of this nature
constitute a derivational approach to L.

An alternative view, which has been popular at different times throughout
the history of generative grammar places the explanatory burden on con-
straints that apply to complete structural descriptions (see especially
McCawley (1968), Jacobson (1974), as well as Chomsky (1981), Bresnan
(1982), Rizzi (1990), and Brody (1995). This “representational” perspective
does not appeal to intermediate structures, and it has proved difficult to dis-
tinguish the two empirically.

However, evidence in favor of a derivational approach has been brought
to bear by Gereon Müller (1998) in a penetrating “remnant movement”
analysis of incomplete category fronting in German. The idea of remnant
movement (Thiersch 1985, den Besten and Webelhuth 1987) is to analyze a
2. Classical analysis

2.1. The phenomenon

It is well-known that German and Dutch allow both topicalization and scrambling. The following data\(^4\), from Müller (2000) illustrate these options. The first is VP topicalization, shown in (4) below.

(3) \[ \text{Vp Das Buch gelesen, hat keiner,} \]

\[ \text{the book read has no-one} \]

"No one read the book"

Example (3) depicts an analysis of topicalization as WH-movement (Chomsky 1977) in which the verb phrase \text{Das Buch gelesen} has landed in the specifier of a functional projection, perhaps CP or TopP\(^5\). The second independently available option is NP scrambling, shown in (4) below.

(4) \[ \text{daß das Buch, keiner Vp, gelesen, hat} \]

\[ \text{that the book no-one read has} \]

"that no one read the book"

Example (4) is an embedded clause headed by the overt complementizer \text{daß}. The NP \text{das Buch} has moved in (4) from its base position as a complement of the verb \text{gelesen} to a position still to the right of \text{C}. Taking for granted that \text{CP} (or \text{TopP}) specifiers come to the left of their heads, the movement of \text{das Buch} is evidently distinct from the one in (3) since (at least) the landing sites differ. Such NP scrambling is standardly analyzed as left-adjunction to \text{IP} or \text{VP} (Müller 1995, chapter 3). Both phenomena can be found in the same sentence, as (2) – repeated here as (5) – demonstrates\(^6\).

(5) \[ \text{Vp Gelesen, hat das Buch, keiner} \]

\[ \text{read has the book no-one} \]

"No one read the book"

In a main clause, additional movement following the scrambling seen in example (4) leaves trace \text{t} \(_{2}\) and is termed "remnant" because the moved constituent itself contains a trace, \text{t} \(_{1}\).

This analysis is attractive because the feeding relation between the two transformations provides cross-linguistic insight. Such a double-movement analysis immediately predicts that languages that do not independently pos-
sness the first kind of movement can never show incomplete category fronting. This prediction is confirmed by the contrast between the scrambling languages German and Dutch, which do display incomplete category fronting, and all other Germanic languages, which are non-scrambling and do not display incomplete category fronting (den Besten and Webelluth 1987, 15). In a completed theory, remnant movement might serve as the empirical basis for a scrambling parameter whose setting is independent of the topicalization parameter.

Despite being attractive from a cross-linguistic perspective, remnant movement analyses raise theory-internal problems for a Principles and Parameters (P&Ps) approach. One problem is that the trace \( t_j \) in (5) is unbound. This contradicts the presumably inviolable Proper Binding Condition (Fieno 1977, 45) which requires that moved elements precede their traces\(^7\). Another problem is the violation of the Frozen Structure Constraint (Ross 1967, 173; Wexler and Culicover 1980, 120) which prohibits constituents from moving after subparts have moved. The topicalized VP in (5) is a counterexample to such a principle. Müller argues that while these considerations are problematic for a representational P&Ps approach, they are natural consequences of a derivational MP analysis. The next section examines some of the facts that underlie this conclusion.

### 2.2. The generalization

Remnant movement can occur with a variety of movement type pairs, but not, it appears, when the two types are the same. Example (6) from (Müller 1996, 360) shows that scrambling a remnant VP, rather than topicalizing it, results in ungrammaticality (cf. example (5)).

(6) *daß [\(v_p\) \(t_j\) Gelesen \(t_j\)], das Buch, keiner \(t_j\) hat that read the book no one has

“As noted earlier, the movement in example (6) must not be topicalization, since the landing site is to the right of the complementizer. The same pattern shows up with other combinations of movement. Example (7) from (Müller 1996, 363, 374) contrasts the acceptability of the scrambling/WH-movement combination (7-a) with the unacceptability of another scrambling/scrambling combination (7-b).

(7) a. [\(n_p\) Welches Buch \(t_j\) hat \(i_p\) über die Liebe \(t_j\), niemand \(t_j\) gelesen? which book has about the love no one read

"Which book about love did no one read?"

b. *daß niemand \(n_p\) ein Buch \(t_j\) gestern \(i_p\) über die Liebe \(t_j\), \(t_j\) that no one a book yesterday about the love

gelesen hat read has

“that no one read a book about love yesterday”

The generalization that emerges in Müller (1996) is stated in (8). Equivalent formulations have been independently proposed by several researchers.\(^8\)

(8) Remnant XPs cannot undergo Y-movement if the antecedent of the unbound trace has also undergone Y-movement.

Müller (1998) undertakes to derive this generalization from MP principles, thereby solving the problems of Proper Binding and Freezing that thwart a GB-type approach.

### 2.3. Derivational account of the basic facts

The essence of the derivational account lies in two principles: Last Resort and the Barriers condition. The Barriers condition primarily addresses Freezing.

(9) Barriers Condition: Movement must not cross a barrier. (Müller 1998, 31)

(10) Barrier: XP is a barrier for \(\alpha\) iff there is an \(X^\alpha\) \((0 \leq n \leq P)\) such that (a), (b) and (c) hold

a. \(X^\alpha\) includes \(\alpha\)

b. \(X^\alpha\) is not a complement

c. \(X^\alpha\) is distinct from \(Y^\alpha\), where XP is the complement of \(Y^\alpha\)

The import of this definition is that moved constituents become barriers when they move to a specifier or adjunct position, since all derived positions...
are specifiers or adjuncts, this implements the Freezing generalization. The
other key principle is Last Resort. Last Resort (Müller 1998, 273) is construed
to include Chomsky’s MLC (1) as part ‘b’ of its definition.

(11) Last Resort: α is raised to a position β only if (a) and (b) hold

(a) β is a typical checking position for the lowest-ranked unchecked
morphological feature F of α

(b) There is no γ with an unchecked F feature that is closer to β

It is the MLC, part b of Last Resort, which is primarily responsible for
explaining the distribution of remnant movement in German. It does this by
ensuring that a subpart of a constituent is never raised if, later in the deriva-
tion, the entire constituent would also raise to check the same kind of feature.9

These two principles interact to yield a derivation of the acceptable
example (2) [schematically repeated here as (12)] but no derivation for the
unacceptable example (6) [repeated here as (13)].

(12) a. …Top[Top] ⋯ Y[scr] ⋯ [α[Top] ⋯ NP[scr] ⋯ ] ⋯ 

b. [VP t₁ Gelesen ] hat das Buch, keiner t₂
read has the book no-one
“No one read the book”

(13) a. …X[scr] ⋯ Y[scr] ⋯ [α[scr] ⋯ NP[scr] ⋯ ] ⋯ 

b. * daß [VP t₁ Gelesen ] das Buch, keiner t₂ hat
that read the book no one has
“that no one has read the book”

In both cases, the inviolable MLC requires that α move to a specifier of the
head with the closest checkable feature (Y). Example (12-b) is grammatical
because it has a derivation that does not violate any principles. In particular,
the lowest NP moves to Y, checking the scrambling feature [scr], then the
remnant α moves to Top, checking [top].

The unacceptable scrambling/scrambling derivation would have to pro-
ceed from a derivational stage containing two of the same features. Consider
some of the most likely scenarios. Part b of Last Resort favors movement of
α to the specifier of Y. But such a movement cannot derive (13-b) because, by
failing to evacuate [Top das Buch], α is precluded from becoming a remnant.
Even if α does move to the specifier of Y, under the Barriers Condition, α
then freezes and becomes a barrier to further movement, preventing [NP das
Buch] from escaping. All possible (feature-driven) movement options result
in a violation, and so the ungrammaticality of (13-b) is obtained. All in all
(13-b) is ungrammatical because it has no legitimate derivations.

2.4. Character of the explanation

Why do Last Resort and the Barriers Condition suffice to explain the general-
ization in (8)? A closer look at the formal principles of this analysis reveals
that the issue of derivationality is surprisingly absent. Rather, the explanation
derives entirely from the link-minimization property expressed locally in
the MLC. Consider an input with two of the same features such as (14) below.

(14) scr1 [scr] ⋯ scr2 [scr] [α[scr] ⋯ β[scr]]

In underlying structure (14) there are two scrambling features which need to
be checked, separated from each other by a number of maximal projections
M, and from the largest potential checker by N maximal projections. Given
(14) as input, the preference for non-remnant output is derived solely on the
basis of link-minimization, where links are measured in terms of the number
of maximal projections spanned.

From an input like (14), the cost to reach the remnant movement configu-
ration is N+1 for β (perhaps a verbal argument) and N+M for α (the verb
phase) for a total of 2N+M+1. To reach a nonremnant configuration in which
α checks scr2 and β checks scr1 costs M+1 for β and N for α, totaling
N+M+1. If N>0, which is to say, if scrambling is analyzed as adjoinment to
VP, then given (14) as input, the nonremnant surface structure necessarily
has shorter links.

Should the input contain two different features, as in (15),

(15) top [scr] ⋯ [α[top] ⋯ β[scr]]

in which only α can check the topic feature [top], then β must move on its
own. A remnant movement structure is then optimal.

From the vantage point of Last Resort, the exception posed by German
remnant movement to Proper Binding and Freezing is explained chiefly by
part b—the MLC.

The explanation is one of feature faithfulness: the nearest phrase (favored
by the MLC) cannot check the relevant feature, so the grammar settles for
feature checking by a phrase marker whose chains are not minimal. The
MLC penalizes chains of greater lengths, some of which nonetheless do figure
in grammatical sentences. Couching this style of explanation in OT terms, the MLC functions as a gradient markedness constraint on chain
lengths which is surface-violated in the case of remnant movement. Evidently, the derivational character of the explanation subsists solely in the
presentation of Last Resort.

3. Non-derivational proposal

The previous section reviewed an MLC-based account of the fact that in complete category fronting is possible when movement types are different,
but apparently impossible when they are the same. The essential insight of
this account was found to be neutral between derivational and representa
tional grammars. So, in principle there is no reason why a notational variant
could not be constructed to cover the same data. However, if the essential
insight is correct, it should extend to other languages. The next section con
siders some further data gathered with a view toward such extension. Then a unified, non-derivational analysis is provided.

3.1. Typological considerations

A remnant is a moved phrase which itself contains a trace of movement. But
what is a non-remnant? This question is really a typological question about
how languages that exhibit remnant movement differ from other languages.
To facilitate this sort of comparison, it will be necessary to fix some dimen
sions of typological variation. For purposes of cross-linguistic comparison,
let the number of movements under consideration be restricted to two. Then the general form of a remnant movement structure is given in (16).

(16) \( \ldots [\alpha \ldots \ t_i \ldots ] \ldots [^M \ldots \beta_{i+1} \ldots [^N \ldots \ t_j \ldots ] \ldots \)

In words, a remnant movement structure (16) is a phrase structure that has some moved phrase \( \alpha \) separated from a moved sub-constituent \( \beta_i \). The launching and landing sites are separated by some number of projections indicated by \( M \) and \( N \).

What is the corresponding non-remnant structure that is most naturally comparable to (16)? An attractive counterpart is (17).

(17) \( \ldots \beta_i \ldots [^M \ldots [\alpha \ldots t_i \ldots ] \ldots [^N \ldots t_j \ldots ] \ldots \)

Structure (17) differs only from (16) in that it eliminates the property of \( \alpha \)
being a remnant. These structures have been named “diving” (16) and “surfing” (17) by Pesetsky (1982). The terminological shift is necessary to make the concept of “non-remnant” movement less vague. More fundamentally,
these two structures provide a kind of ruler for examining cross-linguistic variation.

Applying these rulers, one can immediately observe (2) that German exhibits diving movement when given input containing different features,
such as scrambling and topicalization. Equally immediate is the observation
(6) that German does not permit diving movement when two of the same
features are present.

Do any languages generate surfing configurations from these inputs? Japanese does. In Japanese, surfing is attested in connection with (long-distanc

(18) Booru-o, matigatte [\( \text{CP} \) Susi-ni \( \text{t_j} \) wataasoo to \( \text{I} \) ], Kazuko-ga \( \text{t_k} \) kokoromita
ball\text{ACC} falsely Susi-to to give that Kazuko\text{NOM} tried
"Kazuko tried to give a ball to Susi by mistake"

Namely, when given two scrambling features, the Japanese result is not ungrammaticality, but instead, the surfing movement configuration. Now consider contrasting movement types in Japanese. Here, diving or remnant movement is indeed possible. Example (19) from Tsujioka (2000) citing Kurafuchi (1995) illustrates this for the scrambling/WH movement combination.

(19) \( [\text{CP} \text{[iSC} \text{t_i} \text{donna-ni kirei-ni j}] \ldots [\text{[TP biyooi ga [DP Mary o] t_i sita] no}] \}
how much beautiful beautician NOM Mary ACC did Q
"How beautiful did the beautician make Mary?"

These findings are summarized in Table 1 below.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>different</th>
<th>same</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>diving</td>
<td>*</td>
</tr>
<tr>
<td>Japanese</td>
<td>diving</td>
<td>surfing</td>
</tr>
</tbody>
</table>

Table 1. Patterns of diving and surfing
3.2. An OT approach to remnant movement in German and Japanese

This subsection reanalyses the generalization (8) “remnant XPs cannot undergo Y-movement if the antecedent of the unbound trace has also undergone Y-movement” (Müller 1998, 240) from the perspective of Harmonic Parallelism. After motivating the constraints, the inputs, and the candidate set, the bulk of the observations in Table 1 are straightforwardly explained as a Strict Domination interaction between a preference for short chains and a preference for fully-checked features. Adding in the notion that feature checking itself is violable explains the full range of phenomena in Table 1 and predicts new languages.

3.2.1. Constraints

Section II identified the crucial role of the MLC in Müller’s (1998) grammar of German remnant movement. A similar, but not identical function is served by the OT MinLink constraint family (Legendre et al. 1995, 1998).

\[(20) \text{BAR}^k \text{A chain link must not cross } k \text{ barriers.} \]

[where a barrier is a maximal projection that is not L-marked]

(Chomsky 1986)

\[(21) \text{MinLink (universal): BAR}^a \gg \ldots \gg \text{BAR}^2 \gg \text{BAR}^1 \]

[Local conjunction power hierarchy of BAR]

BAR functions only as a building block for MinLink, which registers one violation for each link of each chain in a candidate. For example, BAR is the Local Conjunction (Smolensky 1997) of BAR with itself, indicated by the symbol &. This conjunction can be recursively applied to derive a markedness constraint against chain links of any length: MinLink.

MinLink is motivated by data from Chinese, Bulgarian, and English. For example, it derives the suboptimality of English superraising.

Tableau 1. English super-raising (Legendre et al 1998)

<table>
<thead>
<tr>
<th>Input: LF(=)seems(likely(win(John)))</th>
<th>BAR(^a)(=)BAR(^1) &amp; BAR(^1)</th>
<th>BAR(^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. It seems [ that [ John, is [ likely [ (t) to [ win]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. John, seems [ that [ it is [ likely [ (t) to [ win]</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Tableau 1 displays the reason why super-raising (candidate b) is ungrammatical in English: shorter chains are preferred to longer ones. Candidate b incurs a violation of BAR\(^2\) compared with a violation of BAR\(^1\) for candidate a. Therefore a is optimal.

That MinLink must be stated as a constraint family – more precisely as a universal hierarchy of individual BAR constraints – rather than a single constraint like the MLC is motivated in Legendre et al. 1998 by the necessity of intercalating faithfulness constraints with the BAR hierarchy. See the referenced article for details.

While MinLink has been applied in contexts where it subsumes much of the role of the Condition on Extraction Domain (CED) by penalizing intermediate traces, the two are not equivalent. In particular, MinLink is crafted with violability in mind whereas the CED was not. However, it appears that the CED, in its violable version is not quite on the right track as freezing evidence from German suggests.

This is evident in counter examples to the Barriers condition defined earlier. This CED-like condition makes explicit the conditions under which the “properly governed” clause of Huang’s original CED actually obtains. If the CED or Barriers condition were inviolably surface-true, then (22) ought to be ill-formed (de Kuthy and Meurers 1999, 5).

\[(22) \text{Worüber kann [einen Südkurier-Artikel] selbst Peter \(nicht\) am Strand verfassen?} \]

[beach write]

“For which topic is it the case that even Peter cannot write an article about it for the Südkurier when he is at the beach?”

Here, assuming the NP einen Südkurier-Artikel has scrambled from object position, the CED or Barriers condition ought to rule out any further extraction at the derived position. But in (22), it appears that the WH-element worüber has been moved in just this way. Another case is where again the object precedes the subject, suggesting scrambling, but additional movement of a different kind is still possible (Fanselow 2000, 14).
In a more complete grammar, such base orderings might be weakly constrained to match canonical information structures, perhaps via “imperfect correspondence” (Bresnan 2000, section 4).

The grammar’s function is to define the inventory of acceptable chains, so it is natural to assume the input also contains “underived” phrase structure, which might affect the chain-formation decision. Characteristic of Harmonic Parallelism, outputs are full structural descriptions of surface forms.

3.2.3. Candidate sets

The grammar specifies Harmony maxima that represent the best-formed chains that are as compatible as possible with the input. For this reason, the relevant aspects of Gen are the ways it defines chains. Since the present analysis adopts MINLINK, candidate sets can be restricted to a maximum chain-length n where n is the number of syntactic features in the input. In the presence of MINLINK, chains longer than n will always be suboptimal: these structures are “harmonically-bounded” by candidates whose chains are shorter.

3.2.4. Tableaux for basic German and Japanese patterns

The analysis is simply that, in German, faithfulness to input features overrides the preference for shorter chains, licensing remnant movement – the diving configuration of (16) which is the structural description for the example (2). This argument, played out in Tableau 2, preserves the insight of Müller (1998, chapter 5.7)[2]. Since two different chains are present and neither of them originate in non-complement position, MINLINK and MINLINK cannot be distinguished – an instance of encapsulated constraints (Prince and Smolensky 1993, chapter 8; Legendre et al. 1998).

Tableau 2. Two different features – German

<table>
<thead>
<tr>
<th>Input: Top</th>
<th>FF</th>
<th>MINLINK</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. das Buch hat keiner</td>
<td>i=N-1, k=N+M</td>
<td></td>
</tr>
<tr>
<td>b. das Buch gelesen hat keiner</td>
<td>k=N+M</td>
<td></td>
</tr>
<tr>
<td>c. Top</td>
<td>**</td>
<td>k=N+1, i=N-1</td>
</tr>
<tr>
<td>d. das Buch, hat gelesen ] keiner</td>
<td>***</td>
<td>k=N+1, i=N-1</td>
</tr>
</tbody>
</table>
Given an input that specifies topicalization and scrambling, in that order, the
optimal candidate (a) has two chains. One chain, headed by das Buch and
having t as its tail, crosses N+1 maximal projections; one for the verb
phrase itself and N as determined by an underlying configurational theory
of German phrase structure. The other chain subscripted k crosses N+M
maximal projections. Eval deals with MINLINK just as any other markedness
constraint, by adding up all violations to produce a single integer number
of stars. For notational convenience we will continue to write pseudo-equations
of the form “chain-index = sum of maximal projection counts” to indicate
where the violations are coming from. Unchecked features, sensitive to FF are
in bold. The last candidate (d) in table 2 suffers two FF violations because
the wrong constituents have checked incompatible features.

In Tableau 2 any alternative to remnant movement requires some form of
feature neutralization, violating FF. Although topicalizing the verb phrase
das Buch gelesen results in a less-marked structure (b) with only one chain
rather than two, Strict Domination of MINLINK by FF rules it out.

The same constraint conflict derives the preference for surfing over
diving in Japanese. This is depicted in Tableau 3 (examples from Sauerland
1996).

Tableau 3. Two of the same features – Japanese

<table>
<thead>
<tr>
<th>Input: Ser [′n Ser [′[c′]i[no]... Booru-o[no]... ] ]</th>
<th>FF</th>
<th>MINLINK</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. eFr Booru-o, matigatte [c Susi-ni teto watoso to], Kazuko-ga t kokoromita</td>
<td>i=N, j=M+1</td>
<td></td>
</tr>
<tr>
<td>b. [c Susi-ni teto watoso to], matigatte Booru-o, Kazuko-ga t kokoromita</td>
<td>i=N+M, j=N+1</td>
<td></td>
</tr>
<tr>
<td>c. Ser Ser matigatte Kazuko-ga [c Susi-ni Booru-o watoso to]</td>
<td>**!</td>
<td></td>
</tr>
</tbody>
</table>

Tableau 3 applies the constraint motivated by German to the case in which
two input features are presented to the Japanese grammar. In such a case, both
surfing (a) and diving structures (b) succeed at canceling all input features,
so MINLINK becomes active. As in German, diving movement is always
more expensive than surfing movement.

Tableau 4 illustrates a third application, again to Japanese, in the case
where an input specifies two different features. As in German, Feature Faith-
fulness is the decisive factor when input features differ.

| Tableau 4. Two different features – Japanese |
|---------------------------------------------|---|
| Input: WH [′n Ser [′n [c′]t[no]... Mary-o[no]... ] ] | FF | MINLINK |
| a. eFr [c Susi-ni teto watoso to], biyooi ga [c Susi-ni Mary-o], teto sita no | i=1, j=1 |
| b. [c Susi-ni Mary-o], biyooi ga [c Susi-ni Mary-o], teto sita no | i=1 |
| c. WH [c Susi-ni Mary-o], biyooi ga [c Susi-ni Mary-o], teto sita no | **! |
| d. [c Susi-ni Mary-o], biyooi ga [c Susi-ni Mary-o], teto sita no | **! |

Across three cells of Table 1, repeated below, a broadly painted picture of
markedness-faithfulness conflict between FF and MinLink suffices to char-
acterize both the German and Japanese situation. The intuition of Müller,
Kitagawa, Takano, Fukui and others has found its place in a Harmonic
Parallelism analysis.

Table 1. (repeated)

<table>
<thead>
<tr>
<th>different</th>
<th>same</th>
</tr>
</thead>
<tbody>
<tr>
<td>German</td>
<td>diving</td>
</tr>
<tr>
<td>Japanese</td>
<td>diving</td>
</tr>
</tbody>
</table>

The remaining German cell, however motivates a re-ranking account of the
difference between German and Japanese.

3.2.5. The difference between German and Japanese

One well-known difference lies in the fact that German scrambling, unlike
Japanese scrambling, cannot cross a finite clause boundary (Grewendorf
1993, 1308). But this is exactly the property which would allow one to
observe either surfing or diving configurations in which the “first” move-
ment is VP scrambling. It is therefore not obvious what sort of structure
corresponds to the Japanese input in Tableau 3.

The appropriate comparison would be to cases in which same-feature-
driven movement is ‘almost’ observable. Indeed there are said to be some
cases where predicative scrambling is marginally acceptable in German
(Müller 1996, 362).
(25) a. *daß [vg dem Peter t₁ gegeben]ₖ die Claudia einen Kuß t₂ hat that ART PeterDAT given ART Claudia a kiss has "that Claudia gave Peter a kiss"

    b. ??daß [vg dem Peter einen Kuß gegeben]ₖ die Claudia gestern t₂ hat that ART PeterDAT a kiss given ART Claudia yesterday has "that yesterday Claudia gave Peter a kiss"

What happens when the German grammar is presented with an input where two features of the same type are presented? Example (25-b) suggests that diving is not the result. But some sort of neutralization is occurring. One possibility is that some scrambling features in German do not delete after checking. Sauerland (1996, 1998) observes that Chomsky's (1995) distinction between interpretable and non-interpretable feature is borne out in the difference between Japanese, where surfing is possible (Tableau 3) and German, where it is not (26).

(26) a. *?weil [den Ball], vergeblich [der Susi t₁ zu geben] die Kazuko t₂
    since the ball unsuccessfully the Susi to give the Kazuko
    versucht hat
    tried has

    b. *?weil gestern [von Chomsky]₁ in Frankfurt [das neue Buch t₂]
    since yesterday of Chomsky in Frankfurt the new book
    niemand t₃ gekauft hat
    nobody bought has

Adopting this key idea offers the possibility that it is the correspondence between check-er and check-ee that is neutralized in this cell of Table 1. In fact, Sauerland's proposal can be stated without appeal to feature strength in OT terms, by splitting FF into a pair of violable constraints conspiring to promote one-to-one feature checking.

(27) *MANYTOONE: Violated if the same +F feature checks multiple -F features

(28) *ZERO TOONE: Violated if a -F feature goes unchecked.

On a parallel correspondence theory of feature checking, *MANYTOONE is familiar from phonology as the two constraints UNIFORMITY and INTEGRITY in which the correspondents are now syntactic features (McCarthy and Prince 1995, appendix A).

In OT terms, Sauerland's finding that +F features are re-checkable in German but not in Japanese amounts to the assertion that *MANYTOONE plays a greater role in Japanese than in German. Trading places, however, with all of MINLINK will not work. Rather, the sensitivity of particular members of the MINLINK family that can discern between complement extraction and noncomplement extraction is called for. A descriptively adequate ranking for German is

(29) German: *ZERO TOONE >> MINLINKNCOMPL >> MINLINKCOMPL >> *MANYTOONE

Tableau 5 suggests that the winning candidate when the German grammar is presented with an input containing two of the same features is a compromise like (25-b) where a single VP checks two scrambling features, avoiding both surfing and diving.

Tableau 5. Two of the same features – German

<table>
<thead>
<tr>
<th>Input: Ser</th>
<th>¹⁰ Scrolling ¹{[vg dem Peter] einen Kuß [von]₉ …} …</th>
<th>*Z-1</th>
<th>MLNCOMPL</th>
<th>MLCOMPL</th>
<th>*M-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ñeßdaß [vg dem Peter einen Kuß gegeben], die Claudia t₁, gestern t₂ hat</td>
<td></td>
<td>j=N+1</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. daß einen Kuß [vg dem Peter] t₁ gegeben], die Claudia t₂ hat</td>
<td></td>
<td>i=N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. daß [vg dem Peter t₁ gegeben], einen Kuß t₂, die Claudia hat</td>
<td></td>
<td>j=M+1</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d. Ser die Claudia Ser gestern [vg dem Peter einen Kuß gegeben] hat</td>
<td></td>
<td>j=N+1</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The winning candidate in Tableau 5 is the partially acceptable predicative scrambling item (25-b). In this item, the VP₉ dem Peter einen Kuß gegeben checks both scrambling features. This is notated with an intermediate trace t₁, suggesting the place where the first instance of checking took place. The overall idea is that German cannot permit multiple scrambling of constituents and their subparts, and would rather bend the "rules" of feature checking to avoid it. This idea is implemented by *Z-1's outranking *M-1: while it is still important to enforce the 'interface' condition that all -F features
must be eliminated, that this be done in a completely uniform way is less important. The explanation of the clause-boundness of German scrambling is that it is always cheaper to violate *M-1 using a single VP-chain than it is to break a subpart out of VP.

For Japanese, a descriptively adequate ranking is

(30) Japanese: *ZERO TOONE >> *MANY TOONE >> MINLINK COMPL >> MINLINK NCPL

Tableau 6 verifies the compatibility of the previously-derived Japanese data with this ranking.

**Tableau 6.** Two of the same features – Japanese

<table>
<thead>
<tr>
<th>Input:</th>
<th>Ser [* Ser[* [CP ... Booru-o ...] Susi-ni ...]]</th>
<th>*Z-1</th>
<th>*M-1</th>
<th>ML-CMPL</th>
<th>ML-NCPL</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>[CP Susi-ni Booru-o watatoo to], matigatte t', Kazuko-ga ti, kokoromita</td>
<td>*!</td>
<td>i=N+M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>Booru-o, matigatte [CP Susi-ni ti watatoo to], Kazuko-ga ti, kokoromita</td>
<td>i=N</td>
<td>i=M+1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>[CP Susi-ni ti watatoo to], matigatte Booru-o, Kazuko-ga ti, kokoromita</td>
<td>i=N+1, j=N+M</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.</td>
<td>Ser Ser matigatte Kauko-ga [CP Susi-ni Booru-o watatoo to]</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These rankings also work for both languages when inputs contain different features. The reason is that such competitions are always and essentially a conflict between *ZERO TOONE and MINLINK COMPL whose relative ranking is the same in both German and Japanese. It is in this respect that the new rankings are instantiations of the broad finding that FF >> MINLINK.

3.2.6. Typological predictions

Assuming that *ZERO TOONE has a fixed, high ranking (perhaps in virtue of being an interface condition) the factorial typology is reduced in size from 24 languages to 6, two of which (German and Japanese) have been analyzed in this section. This leaves four other predicted languages.

1. MINLINK COMPL >> MINLINK NCPL >> *MANY TOONE
2. MINLINK COMPL >> *MANY TOONE >> MINLINK NCPL
3. MINLINK NCPL >> *MANY TOONE >> MINLINK COMPL
4. *MANY TOONE >> MINLINK NCPL >> MINLINK COMPL

The input containing two identical features discriminates these predicted languages. Languages 1 and 2 pattern with German, opting for many-to-one feature checking. Language 3 and 4 predict a diving movement path when presented with this input – a prediction which diverges from the Minimalist approach of section II. The larger implication is that the generalization (Takano 2000)

In a derivation yielding the configuration

...[A...t...][...B...t...]

movement of A and movement of B may not be of the same type.

may in fact be language-particular.

4. Empirical basis

The analysis in section III is predicated on the generalization (8) repeated here as (31) which has been widely assumed in the literature.

(31) Remnant XPs cannot undergo Y-movement if the antecedent of the unbound trace has also undergone Y-movement.

However, survey data collected by Gisbert Fanselow indicates that the situation is somewhat more complex. Fanselow finds that, indeed, predicative VP scrambling is possible under some conditions, as illustrated in (32) and (33) below.

(32) dass man ja [vp sein Essen sich zusammenstehlen], nur that one one's food REFL steal-together only [pp in höchster] ti Not darf in greatest need is-allowed

(33) dass [vp die Maria geküsst], dann DOCH keiner ti hatte that the Maria kissed then PTC no-one has
What's more, with the correct intonation, items such as (34) are acceptable.\textsuperscript{15}

(34) dass [vp t, geküsst] [np die Marie] dann DOCH keiner t, hatte that kissed the Marie then PTC nobody had

Sentence (34) involves a scrambled VP containing a scrambling trace. It is a counterexample to generalization (8) where Y is scrambling. A precise characterization of the factors that license exceptions to generalization (8) (such as intonation, the particle doch et cetera) remains an important challenge to any descripively adequate account of German word order.

On the view expounded in section III, though, the situation is quite clear: sometimes German is Japanese, as regards recursive scrambling. In OT terms, sometimes the *MANYTOONE constraint can outrank (both instantiations of) MINLINK. If this does indeed happen in the idiolect of a single speaker, it can be viewed as a case of floating constraints (Reynolds 1994, Antilla 1996, Legendre et al. 2002).

5. Comparison

Any analysis of the difference between German and Japanese will have to assign some formal element that differs across those two languages, be it a lexical element, additional derivational complexity or a parameter, more or less abstract that is set differently across languages. The proposal of section III is that only the relative priorities of universally-given principles are different across German and Japanese. At present it is unclear what mechanism for cross-linguistic diversity will replace parameter setting in the Minimalist program\textsuperscript{16}.

By contrast, the principles required in the new proposal are well-established and very basic: a notion of barrier sensitive to the property of being a complement, and a pair of constraints that tend towards the notion of one-to-one feature checking that has always been assumed in discussions of feature-driven movement – but are crucially violable.

Occasionally, in discussions of OT syntax, the issue of computational complexity is brought up\textsuperscript{17}. Indeed, this general issue is as central for theoretical linguistics as it is for natural language processing (see, e.g. Rounds 1991). However, mention of it is often prompted by a concern that because the set of candidates specified by Gen is infinite, the Harmonic Parallelism view of OT is unworkable – or at least, unimplementable in the brain or some other variety of computer.

Given that generative grammar has always dealt with infinite sets (through recursion) this is a curious criticism. We feel it is out of place in the context of a competence grammar\textsuperscript{8} whose role is taken to be the specification of the set of grammatical expressions in an explicit way, analogous to the notion 'generate' in algebra.

To the extent that the competence/performance distinction allows work on the specification of the human language faculty to proceed independently from work on implemented models of that faculty (e.g. Hale and Smolensky 2001) it is our own feeling that theoretical linguistics can constructively engage other cognitive sciences while still remaining insulated from the algorithmic exigencies of search space, temporal sequencing and memory.

6. Conclusion

The main results of this study, however, are purely grammatical. An investigation of the role of the MLC in accounting for German incomplete category fronting (Müller 1998) challenges a prominent argument for derivational grammar. In fact, the key insight of a remnant analysis turns out to be one whose expression in a non-derivational OT framework is quite natural: FEATURE FAITHFULNESS dominates MINLINK. A cross-linguistic comparison of German with Japanese confirms the violable nature of feature checking. This leads to an account of remnant movement in German which – rather than being hindered by violability – is made possible by it. To the extent that the proposed Harmonic Parallelism analysis is successful, it vitiates one of the few empirical arguments for derivational grammar, and supports the view of the MLC as a global, rather than local optimization principle.

Acknowledgements

The authors wish to thank Ralf Vogel, Artur Stepanov, Fabian Heck, Yuji Takano, Colin Wilson, Paul Hagstrom, Luigi Burzio and Paul Smolensky for discussion that has been valuable in the preparation of this paper. We also thank both Florian Wolff and Julia Hockenmaier for German acceptability judgments and valuable discussion. Special thanks to Gisbert Fanselow for his survey results, and to our anonymous reviewers.
Notes

1. Non-remnant movement analyses have also been proposed which appeal to reconstruction (e.g. Grewendorf and Sternefeld 1990) or deny the existence of scrambling on which remnant movement rests (Fanselow 2000). We abstract away from these alternatives because the primary issue of interest here is to compare current derivational and representational approaches and not to compare the merits of all possible accounts of incomplete category fronting in German.

2. “On a more general note, it has often been observed that representational analyses of syntactic phenomena can usually be rephrased in derivational terms (and vice versa) without too much difficulty, so most syntactic phenomena do not force a decision as to what the overall organization of a grammar (representational or derivational) should look like. Therefore, as soon as we find a phenomenon that strongly suggests preferring one approach over the other, this result may have important repercussions that go beyond the question of an empirically adequate account of the phenomenon itself.

   Accordingly, I would like to contend that, on the conceptual side, the analysis developed here should be viewed as an argument for a type of grammar in which all constraints are either derivational or transderivational (and constraints that can be viewed as filters on surface representations do not exist at all).” (Müller, 1998, 323)

3. Müller and his collaborators have recently proposed an OT analysis of remnant movement but it crucially relies on a MP-style derivational approach. The conclusions we reach carry over to those analyses.

4. Müller’s data is reassessed in section IV.

5. We follow Vikner (2001, chapter 3) in assuming the auxiliary hat is in C.

6. Examples of remnant movement analyses in other languages, such as English and Hungarian, can be found in Kayne (1994) and Koopman and Szabolcsi (1998).

7. We follow the original formulation of Proper Binding in terms of precedence (rather than c-command) for expository convenience.

8. Takano (2000, 144) remarks that “Fukui 97, Kitahara 94,97, Müller 96 and Takano 94 have independently proposed to account for the differences between licit and illicit remnant movement by paying attention to the types of movement that affect remnant movement. The basic intuition that we attempt to capture is this: In a derivation yielding the configuration ...[A]...[B]... movement of A and movement of B may not be of the same type.”

9. Müller (1998, 276) puts it this way regarding the configuration: ...B...[A]...[B]... "The crucial observation now is that in this context A is invariably closer to B (in terms of length of path) than x is, the simple reason for this being that B is followed by A."  

10. The appropriateness of separate Scrambling, Topicalization and WH-movement rules for Japanese is a matter of some controversy, and distinguishing them is an active area of research. We are sympathetic with the position of Takahashi (1993, 664) that at least some kinds of long-distance scrambling qualify as WH-movement since they exhibit superiority effects

   (i) John-ga dare-ni [Mary-ga nani-o tabeta to] itta no?
   NOM who DAT what-ACC speaks COMP told Q
   Who did John tell that Mary ate what?

   (ii) Nani-o John-ga dare-ni [Mary-ga ti tabeta to] itta no?
   What did John tell who that Mary ate?

   and have fewer scope possibilities than other kinds of dislocation.

11. One might well begin from a violable constraint like Vikner’s (2002) Proper-Binding, perhaps adding *FROZEN. The goal of the analysis in section III is to provide a satisfactory definition for such constraints.

12. Feature unfaithfulness is but one possible account of these phenomena. Others, perhaps based on bi-directional optimization (cf. Vogel, this volume) may offer similar cross-linguistic insight.

13. The Harmonic Parallelism approach, unlike the Clash and Crash architecture of Broekhuis (2001) does not presuppose that Gen incorporates any economy principles; instead, the MLC resides entirely in the constraint component Con.

14. A native speaker of German informs us that items such as (25-b) may be more acceptable in dialects originating in what was formerly Bohemia. Taking (i) as a base,

   (i) Petr dnes nebude [VP libat Klaudii ruku]
   “Peter will not kiss Claudia’s hand today”

   the VP can be fronted:

   (ii) [VP libat Klaudii ruku], Petr dnes nebude t

   and the NP argument of VP can be scrambled out:

   (iii) Petr dnes [NP Klaudii ruku], nebude [VP libat t]

   but the VP-fronting of (i) is preferable to the double-movement in (iv)

   (iv) * [VP libat t], Petr dnes [NP Klaudii ruku], nebude t

   All four items mean essentially the same thing, so it is reasonable to assume they have the same inputs. Thus, Czech presents a case in which a single movement appears to neutralize two features in a way that two separate movements cannot.

15. Item (34) was judged acceptable by 23 native speakers out of 31 who had an opinion.

16. Müller (1998) remarks “a bit more would have to be said about Japanese” (fn p252).

17. While the complexity of both parsing and generation has been a popular topic in OT phonology, correspondingly less work has pursued these questions in OT syntax. Notable exceptions include Warten (2000) and Kuhn (2001).
18. Chomsky (1968) writes:

"In general, a set of rules that recursively define an infinite set of objects may be said to generate this set. Thus a set of axioms and rules of inference for arithmetic may be said to generate a set of proofs and a set of theorems of arithmetic (last lines of proofs). Similarly a (generative) grammar may be said to generate a set of structural descriptions, each of which, ideally, incorporates deep structure, a surface structure, a semantic interpretation (of the deep structure) and a phonetic interpretation (of the surface structure)."

(footnote 12, page 126).

References

Antilla, A.

Bresnan, J. (ed.)
1982 The Mental Representation of Grammatical Relations. MIT Press.

Bresnan, J.

Broekhuis, H.

Chomsky, N.
1999 Minimalist Inquiries: The framework. Ms, MIT.

de Kuthy, K. and Meurers D.

den Besten, H. and Webelhuth, G.

Fanselow, G.

Fiengo, R.

Grewendorf, G.

Grimshaw, J.

Hale, J. and Smolensky P.

Jacobson, P.

Kayne, R.

Koopman, H. and Szabolcsi A.

Kuhn, J.

Kurafuji, T.

Legendre, G., Hagoort P., Vainikka A. and Todorova, M.

Legendre, G., J. Grimshaw, & S. Vikner (eds.)

Legendre, G., Smolensky, P. and Wilson C.

McCarthy and Prince.
McCawley, J.

Müller, G.

Pesetsky, D.

Prince, A. and Smolensky, P.

Reynolds, B.

Ross, J.
1967 Constraints on variables in syntax. Ph.D. dissertation, MIT.

Rounds, W. C.

Sauerland, U.

Smolensky, P.

Takano, Y.

Thiersch, C.
1985 VP and scrambling in the German Mittelfeld. Ms, University of Tilburg.

Tsujioka, T.

Vikner, S.

Wartena, C.