Local vs. global optimization in syntax: a case study

Gereon Müller
IDS Mannheim

Abstract. The main goal of this paper is to argue for an approach to optimization in syntax that is not global (as is standardly assumed), but local, in the sense that syntactic optimization procedures can affect only small portions of syntactic structure. Local optimization presupposes harmonic serialism (rather than harmonic parallelism), i.e., a derivational organization of grammar. In line with this, I set out to reconcile optimality theory with the minimalist program (see Chomsky (2000), Chomsky (2001)), a derivational approach in which phrase structure is created incrementally. I argue that local optimization is both conceptually attractive (because it significantly reduces complexity) and supported by empirical evidence. As a case study, I develop an analysis of a shape conservation phenomenon in German that involves repair-driven movement operations at the clause edge. I show that, other things being equal, local optimization succeeds where global optimization fails.

1. Background

Optimization can be parallel or serial, and it can be global or local. Optimization is parallel if it only applies once; it is serial if it applies more than once. Following Prince and Smolensky (1993), it is standardly assumed in optimality-theoretic phonology that optimization is parallel.‡ In syntax, too, optimization is usually viewed as parallel.§

The issue of local vs. global optimization has so far received much less attention. An optimization is global if it affects the entire structure of a linguistic expression (e.g., word or sentence); it is local if it applies to a subpart of a linguistic expression. Most of the work in optimality theory relies on global optimization. This is particularly obvious in phonology, but it is also the case in syntax. However, local optimization in syntax is suggested as a possiblity in Archangeli and Langendoen (1997, 214), and in a footnote in Ackema and Neeleman (1998, 478). Full-fledged analyses involving local optimization in syntax include Heck and Müller (2000a), Heck and Müller (2000b), Müller (2000), Fanselow and Ćavar (2001), Heck (2001a), Fischer (2002), and Müller (2002).

Whereas a global approach can be either parallel or serial, a local approach must be serial, such that parts of sentences are successively subject to optimization. In what follows, I sketch a local optimization approach that incorporates main features of the minimalist program, whose incremental-derivational architecture makes it inherently serial.‖

‡ However, see McCarthy (2000), Rubach (2000), and the contributions in Hermans and van Oostendorp (2000) for (discussions of) serial optimization in phonology.

§ See Grimshaw (1997), Pesetsky (1998), Legendre, Smolensky and Wilson (1998), Bresnan (2001), and most of the contributions in Barbosa et al. (1998), Legendre, Grimshaw and Vikner (1998), and Sells (2001). Exceptions that involve serial optimization are typically concerned with syntax/semantics interface phenomena (see, e.g., Heck (2001b) and Hendriks and de Hoop (2001)), and include various systems of bidirectional optimization (see Wilson (2001), Blutner (2000), Jäger and Blutner (2000), Aissen (2002), Lee (2001), Vogel (2002), Jäger (2002)). However, the number of optimization procedures required in these serial approaches is rather small (either 2 or 3).

‖ Pesetsky (1998) and Broekhuis (2000) also combine assumptions of the minimalist program and optimality

2. Approach

Assume that syntactic structure is created incrementally from bottom to top as a result of derivational operations like Merge and Move that have access to the numeration (an array of items selected from the lexicon before the derivation starts). These operations belong to Gen, which also contains inviolable constraints, among them the Strict Cycle Condition (SCC) (Chomsky (1973), Chomsky (2001), Perlmutter and Soames (1979)) and the Phase Impenetrability Condition (PIC) (Chomsky (2000), Chomsky (2001)).

(1) *Strict Cycle Condition* (SCC):
Within the current XP α, an operation may not target a position that is included within another XP β dominated by α.

(2) *Phase Impenetrability Condition* (PIC):
The domain of a head X of a phase XP is not accessible to operations outside XP; only X and its specifier(s) are accessible to such operations.

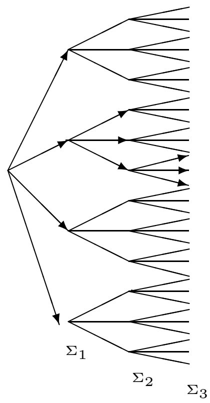
Focussing on Move operations here and in what follows, the SCC restricts the space in which the derivation can find a landing site for movement (i.e., locate the "probe"), whereas the PIC restricts the domain in which the derivation can find an item to move (i.e., a "goal"). The local domains are not completely identical: Every XP is a cyclic domain for the SCC, but only those XPs that qualify as phases (which for present purposes I assume to be CPs) are domains for the PIC.

Where, then, does optimization enter the picture? The idea is that certain derivational units act as local optimization domains Σ, and that Gen and H-Eval apply as many times as there are Σs in the derivation. More specifically, suppose that on the basis of one and the same input, syntactic operations (Merge, Move, etc.) can apply in accordance with inviolable constraints (SCC, PIC, etc.) in different ways, yielding different outputs at stage Σ_i. These outputs are then subject to optimization along a set of ranked and violable constraints, and the optimal output is determined. Only an optimal output can show up in the input of subsequent derivational steps (together with items taken from the numeration and other optimal outputs), and the derivation proceeds in various Gen-compatible ways, producing different outputs at the next optimization domain Σ_{i+1}. At this point, optimization starts anew, yielding a winning candidate that acts as part of the new input, and so on, until all material of the numeration is used up, the derivation reaches an end, and the optimal root clause is determined. Importantly, all locally suboptimal outputs are disregarded in subsequent derivational steps. Therefore, local optimization significantly reduces complexity, compared with global optimization. This is shown schematically in figure 1.

Here, arrowed lines correspond to sequences of derivational steps that yield outputs which participate in local optimization. The other lines represent continuations of suboptimal outputs that give rise to many more outputs. These latter continuations and their associated outputs are simply not available in the local optimization approach adopted here; however, they must also be considered in a global approach, which cannot discriminate between arrowed and other lines. Consequently, a global approach is inherently more complex than a local approach. (It is worth emphasizing that this consequence arises in all global optimization approaches, independently of whether Gen is derivational or representational – in the latter case, the non-arrowed lines encode locally suboptimal subtrees.) Clearly, the degree to which local optimization and global optimization differ with respect to complexity depends on the choice of optimization domain: The smaller Σ is, the more local optimization pays off from

theory, but in a less far-reaching way that basically restricts H-Eval to PF-realization and relies on standard – parallel, global – optimization.

Figure 1: The size of candidate sets under local vs. global optimization



the point of view of complexity (Σ = root clause yields a global approach); on the other hand, an extremely small Σ brings with it the danger of leaving hardly any room for optimization (e.g., if Σ = derivational step). I assume that Σ = XP.

Another complexity-related issue arises if we assume serial optimization of growing subtrees: Does an output carry with it old (but, by definition, non-fatal) violation marks incurred by optimal outputs that are embedded in it, or are such violations invisible? It is not clear whether there is empirical evidence that would distinguish between the two options; but given the overall goal of reducing complexity, it seems preferable to assume that only those parts of an output are visible to H-Eval that are accessible in accordance with the PIC. Thus, only the structure from the present XP down to specifiers of minimally embedded CPs (if there are such) is subject to optimization at any given stage. More generally, we end up with the result that both the *number* and the *size* of competing syntactic outputs are considerably smaller than in systems that employ global optimization; taken together, these steps have the effect of bringing optimization in syntax closer to optimization in (non-phrasal) phonology and morphology, where such complexity issues are much less worrysome to begin with (essentially because words are smaller objects than sentences).

I take conceptual considerations like these to be suggestive; but eventually, the question of local vs. global optimization in syntax must be decided on the basis of empirical evidence. To this end, I provide an empirical argument for local optimization. The structure of the argument is as follows: (i) There is evidence for repair-driven movement at the edge of German clauses. (ii) Repair operations strongly suggest an underlying optimization procedure. (iii) The repair operation does not apply in all contexts in which the ranked constraints would seem to force it. (iv) The contexts in which it does not apply even though the constraints seem to demand application correspond to non-arrowed lines in figure 1 which are irrelevant in local optimization, but must be considered in global optimization.

The evidence I want to discuss involves a well-known asymmetry that shows up with *wh*-movement from embedded clauses in German.

3. Data

Two types of finite declarative clauses can be embedded under bridge verbs in German: (i) clauses headed by a complementizer *dass* (‘that’); (ii) V/2 clauses with finite V in the C position and some XP in SpecC. Both types of complements as such appear to be transparent for *wh*-movement to SpecC. *Wh*-movement from a *dass* clause may go to a *dass* clause or to a V/2 clause; see (3-ab).¶ In contrast, as shown in (3-cd), *wh*-movement from a V/2 clause may only end up in a V/2 clause again (see Tappe (1981), Haider (1984), Reis (1985)).

(3) a. (Ich weiß nicht) [$_{CP_1}$ wen$_i$ (dass) du meinst [$_{CP_2}$ t$'_i$ dass sie t$_i$ getroffen hat]]
I know not whom that you think that she met has

b. [$_{CP_1}$ Wen$_i$ meinst du [$_{CP_2}$ t$'_i$ dass sie t$_i$ getroffen hat]] ?
whom think you that she met has

c. [$_{CP_1}$ Wen$_i$ meinst du [$_{CP_2}$ t$'_i$ hat sie t$_i$ getroffen]] ?
whom think you has she met

d. *(Ich weiß nicht) [$_{CP_1}$ wen$_i$ (dass) du meinst [$_{CP_2}$ t$'_i$ hat sie t$_i$ getroffen]]
I know not whom that you think has she met

The same restriction holds when movement from SpecV/2 to Spec*dass* is followed by further *wh*-movement, or when the moved item is a topic or relative pronoun (the analysis below could be extended in obvious ways to cover topicalization and relativization). The data have proven remarkably robust over the years, and many attempts have been made to account for the asymmetry involved. First, it has been suggested that a V/2 clause acts as an island in (3-d), which then requires some extra assumption about (3-c), where islandhood seems to be voided (see Staudacher (1990), Sternefeld (1989), Reis (1996)). Second, it has been proposed that the asymmetry in (3) follows from directionality constraints on movement (see Müller (1989), Haider (1993)). Third, the data have been approached in terms of constraints against improper movement (see Haider (1984), Sternefeld (1992), Müller and Sternefeld (1993), Williams (2003)). However, all these approaches can be shown to involve construction-specific assumptions, and it seems fair to conclude that the problem in (3) has not yet received a satisfying solution.

4. Analysis

Suppose that movement is triggered by certain types of features on the probe that must be matched by appropriate features on the goal; following Sternefeld (2003), I refer to the features that trigger movement as [*F*] (i.e., "strong") features, with matching [F] features on the goal. Two violable and ranked constraints play a role in this context: FC (*Feature Condition*) ensures that [*F*] on some lexical item X triggers movement to the edge of an XP (the edge of an XP comprises X and SpecX; see Chomsky (2000), Chomsky (2001)). LR (*Last Resort*) requires that movement results in feature matching.

(4) a. *Feature Condition* (FC):
An [*F*] feature on X requires an item bearing [F] at the edge of XP.

b. *Last Resort* (LR):
Movement requires matching of [F] and [*F*] at an edge.

Two further constraints of H-Eval are OP (*Operators at Clause Edges*; based on Grimshaw (1997)) and, crucially, SCE (*Shape Conservation for Clause Edges*). Shape Conservation has been suggested as a general constraint by Williams (2003).[+] Versions of this constraint

¶ A complementizer of CP$_1$ must then be deleted in Standard German, but not in dialects and colloquial varieties. Following Pesetsky (1998), I assume that complementizer deletion is a PF phenomenon in languages like German and English, with a *that*/*dass* complementizer present in syntax proper.

[+] Note that Williams (2003, 78-79) actually provides an account of the pattern in (3). However, Williams' analysis does in fact not rely on Shape Conservation; rather, it is an account in terms of improper movement that is very similar to the approach in Sternefeld (1992) – which in turn can be shown to be based on concepts that are adopted within an optimality-theoretic approach in Müller (2001) (for co-argument NPs) and in Müller (2000) (for VPs). Thus, Shape Conservation can be viewed as a familiy of constraints, of which SCE is a member.

(5) a. *Operators at Clause Edges* (OP):
An operator must be at the edge of a clause.
b. *Shape Conservation for Clause Edges* (SCE):
Clause edges have identical shapes.

SCE is a gradual constraint. Given the edge of a CP_α, SCE violations for CP_β are computed as follows: (i) Compare the n-th edge constituent of CP_α with the n-th edge constituent of CP_β and assign a * if the two items do not have an identical shape. (ii) For each edge constituent of one CP that does not correspond to an edge constituent of the other CP, assign a *.

Assume now a ranking FC $\gg$ OP, SCE $\gg$ LR. On this basis, let me first briefly address successive-cyclic *wh*-movement in general. Unbounded dependencies can be divided intro three parts: a bottom, a middle, and a top (Gazdar et al. (1985)); see (6).

(6) $\underbrace{[_{\mathrm{CP_1}}\mathrm{wh_i}\ \mathrm{C_{[*wh*]}} ...}_{\text{top}}\ \underbrace{[_{\mathrm{CP_2}}\mathrm{t''_i}\ \mathrm{C}\ ...}_{\text{middle}}\ \underbrace{[_{\mathrm{CP_3}}\mathrm{t'_i}\ \mathrm{C}\ ...\ \mathrm{t_i}\ ...]]]}_{\text{bottom}}$

Movement at the top is triggered by FC, given a feature [*wh*] on interrogative C and a matching feature [wh] on a *wh*-phrase. In contrast, movement at the bottom and in the middle is not feature-driven (such intermediate movement steps are required theory-internally by the PIC; they are empirically supported by the existence of visible reflexes of successive cyclicity in the C domain in various languages). Movement that is not feature-driven violates LR; it qualifies as “repair-driven” in the terminology of Heck and Müller (2000a), i.e., it must be forced by a higher-ranked constraint. Movement at the bottom is triggered by OP, given the ranking OP $\gg$ LR (an “operator” in the sense of (5-a) is an XP that bears a feature like [wh]).

Finally, movement it the middle (a notorious problem in incremental-derivational approaches to syntax) is triggered by SCE, given the ranking SCE $\gg$ LR.* Here is why: Suppose that an $XP_{[wh]}$-C shape has been created at the CP_α edge at the bottom. Then, SCE demands a replication of this shape at the next CP_β edge. As long as no higher-ranked constraint precludes this, the SCE thus triggers movement steps in the middle, in violation of LR.♯ At the top, the demands imposed by SCE and FC converge. The question arises as to why SCE does not force *wh*-movement beyond a [*wh*] target position (see Pullum (1979, 372)). This follows from the ranking FC $\gg$ SCE: FC not only forces *wh*-movement to $SpecC_{[*wh*]}$; it also demands that the *wh*-phrase stays in this position.††

Let me now turn to the specific situation in German. Suppose that [*F*] features that can be on C include [*xp*] (for movement of some XP to SpecC), [*wh*] (for *wh*-movement), and [*fin*] (for V/2-movement to C); these assumptions are virtually unavoidable in a feature-based approach to movement. Minimally, there must be two C elements in the lexicon for declarative clauses; these are rendered here as C_d and C_e: C_d = [$_\mathrm{C}$ dass]; C_d does not trigger

were first suggested in Williams (1974).

* OP cannot force movement in the middle because it is satisfied once and for all when the *wh*-phrase has reached the first edge of a clause; see Fanselow and Ćavar (2001), who make use of this property of OP in their account of partial *wh*-movement constructions.

♯ Isn’t an $XP_{[wh]}$-C shape of CP_α destroyed if $XP_{[wh]}$ moves to CP_β? This issue does not arise if traces count for shape conservation. Alternatively, we can conceive of the shape of a CP edge as something that is fixed once and for all as soon as the CP has been optimized.

††What about constructions in which NP-movement to subject position feeds *wh*-movement to SpecC? In these cases, there is no way to avoid a FC violation, and the decision then falls to independent constraints.

any movement via FC. C_e = [$_C$ Ø$_{[*xp*],[*fin*]}$]; C_e triggers V/2 and XP-movement to SpecC. Similarly, there are two C elements for interrogative clauses: C_{dw} = [$_C$ dass$_{[*wh*]}$]; C_{dw} attracts a *wh*-phrase via FC (and is PF-deleted in Standard German). C_{ew} = [$_C$ Ø$_{[*wh*],[*fin*]}$]; C_{ew} triggers *wh*-movement and V/2.

We can now derive the pattern in (3) on the basis of SCE. The two relevant local optimization procedures involve first the embedded CP_2, and then the matrix CP_1. SCE is always vacuously fulfilled in the first optimization procedure, and the optimal CP_2 will either be a *dass* clause or a V/2 clause, depending on the [*F*] features of C . The competition in T_1 is based on an initial choice of C_d that is merged with the optimal TP created by the derivation so far; it produces an embedded *dass* clause as the optimal output, viz., O_2; only this output can then serve as an input for further operations. (Throughout, only the most relevant candidates are shown in tableaux.)

T_1: 'dass' in CP_2: (3-a), (3-b)

Input: [$_{C_d}$ dass], [$_{TP}$ sie wen getroffen hat]	FC	OP	SCE	LR
O_1: [$_{CP_2}$ [$_C$ dass] [$_{TP}$ sie wen getroffen hat]]		*!		
☞O_2: [$_{CP_2}$ wen$_i$ [$_C$ dass] [$_{TP}$ sie t_i getroffen hat]]				*
O_3: [$_{CP_2}$ wen$_i$ [$_C$ hat$_j$ (dass)] [$_{TP}$ sie t_i getroffen t_j]]				**!

The derivation proceeds by optimizing the matrix VP and the matrix TP. Subsequent optimization of CP_1 may then lead to a *dass* clause or to a V/2 clause, depending on the nature of C as C_{dw} or C_{ew}; see T_2, T_3. Consequently, (3-a) and (3-b) are both optimal. However, whereas optimal O_{22} in T_2 respects both FC and SCE (the clause edges have an identical $XP_{[wh]}$-*dass* shape), optimal O_{24} in T_3 must violate SCE by applying V/2 in order to satisfy FC (for [*fin*]): *dass* is in C_2, V/2 is in C_1.†

T_2: Wh-movement from 'dass' clauses into 'dass' clauses: (3-a)

Input: [$_{C_{dw}}$ dass$_{[*wh*]}$], [$_{TP}$ du meinst [$_{CP_2}$ wen$_i$ [$_C$ dass] [$_{TP}$ sie t_i getroffen hat]]]	FC	OP	SCE	LR
O_{21}: [$_{CP_1}$ [$_{C_{dw}}$ dass] [$_{TP}$ du meinst [$_{CP_2}$ wen$_i$ [$_C$ dass] [$_{TP}$ sie t_i getroffen hat]]]]	*!		**	
☞O_{22}: [$_{CP_1}$ wen$_i$ [$_{C_{dw}}$ dass] [$_{TP}$ du meinst [$_{CP_2}$ t'_i [$_C$ dass] [$_{TP}$ sie t_i getroffen hat]]]]				

T_3: Wh-movement from 'dass' clauses into V/2 clauses: (3-b)

Input: [$_{C_{ew}}$ Ø$_{[*wh*],[*fin*]}$], [$_{TP}$ du meinst [$_{CP_2}$ wen$_i$ [$_C$ dass] [$_{TP}$ sie t_i getroffen hat]]]	FC	OP	SCE	LR
O_{21}: [$_{CP_1}$ [$_{C_{ew}}$ Ø] [$_{TP}$ du meinst [$_{CP_2}$ wen$_i$ [$_C$ dass] [$_{TP}$ sie t_i getroffen hat]]]]	*!*		**	
O_{22}: [$_{CP_1}$ wen$_i$ [$_{C_{ew}}$ Ø] [$_{TP}$ du meinst [$_{CP_2}$ t'_i [$_C$ dass] [$_{TP}$ sie t_i getroffen hat]]]]	*!			
O_{23}: [$_{CP_1}$ [$_{C_{ew}}$ meinst$_j$ Ø] [$_{TP}$ du t_j [$_{CP_2}$ wen$_i$ [$_C$ dass] [$_{TP}$ sie t_i getroffen hat]]]]	*!		**	
☞O_{24}: [$_{CP_1}$ wen$_i$ [$_{C_{ew}}$ meinst$_j$ Ø] [$_{TP}$ du t_j [$_{CP_2}$ t'_i [$_C$ dass] [$_{TP}$ sie t_i getroffen hat]]]]			*	

Consider now the case where the optimal embedded CP_2 is a V/2 clause, as in T_4, which uses a different C_2 from T_1, viz., C_e.

† Two remarks. First, the outputs are numbered O_{21}, O_{22}, ... so as to indicate that they are all descendants of O_2 in T_1. Second, O_{22} in T_3 is here assumed to fully respect SCE; i.e., [$_C$ Ø] and [$_C$ *dass*] are taken to have identical shapes (as non-branching C items), in contrast to branching C items that result from V/2. However, this assumption is not crucial; a SCE violation in O_{22} in T_3 would not affect the outcome.

T_4: V/2 in CP_2: (3-c), (3-d)

Input: $[_{C_e}$ $Ø_{[*xp*],[*fin*]}$], $[_{TP}$ sie wen getroffen hat]	FC	OP	SCE	LR
O_1: $[_{CP_2}$ $[_{C_e}$ Ø] $[_{TP}$ sie wen getroffen hat]]	*!*	*		
O_2: $[_{CP_2}$ wen_i $[_{C_e}$ Ø] $[_{TP}$ sie t_i getroffen hat]]	*!			
O_3: $[_{CP_2}$ $[_{C}$ hat_j Ø] $[_{TP}$ sie wen_i getroffen t_j]]	*!	*		
☞O_4: $[_{CP_2}$ wen_i $[_{C_e}$ hat_j Ø] $[_{TP}$ sie t_i getroffen t_j]]				
O_5: $[_{CP_2}$ sie_k $[_{C}$ hat_j Ø] $[_{TP}$ t_k wen_i getroffen t_j]]		*!		

In this case, different choice of C_1 does *not* yield two different optimal outputs in CP_1 optimization. If C_1 has a [*fin*] feature, the optimal CP_1 is also a V/2 clause because of FC, and SCE is respected; see T_5. However, if C_1 does not have such a feature, V/2 will have to apply nonetheless – forced not by FC, but by SCE, in violation of LR; see O_{43} vs. O_{42} in T_6. Here, we have an instance of repair-driven V/2 movement that gives rise to a neutralization effect. This derives the contrast between (3-c) and (3-d); the latter cannot be optimal.‡

T_5: Wh-movement from V/2 clauses into V/2 clauses: (3-c)

Input: $[_{C_{ew}}$ $Ø_{[*wh*],[*fin*]}$], $[_{TP}$ du meinst $[_{CP_2}$ wen_i $[_{C_e}$ hat_j Ø] $[_{TP}$ sie t_i getroffen t_j]]]	FC	OP	SCE	LR
O_{41}: $[_{CP_1}$ $[_{C_{ew}}$ Ø] $[_{TP}$ du meinst $[_{CP_2}$ wen_i $[_{C_e}$ hat_j Ø] $[_{TP}$ sie t_i getroffen t_j]]]]	*!*		**	
O_{42}: $[_{CP_1}$ wen_i $[_{C_{ew}}$ Ø] $[_{TP}$ du meinst $[_{CP_2}$ t_i' $[_{C_e}$ hat_j Ø] $[_{TP}$ sie t_i getroffen t_j]]]]	*!		*	
O_{43}: $[_{CP_1}$ $[_{C_{ew}}$ $meinst_l$ Ø] $[_{TP}$ du t_l $[_{CP_2}$ wen_i $[_{C_e}$ hat_j Ø] $[_{TP}$ sie t_i getroffen t_j]]]]	*!		**	
☞O_{44}: $[_{CP_1}$ wen_i $[_{C_{ew}}$ $meinst_l$ Ø] $[_{TP}$ du t_l $[_{CP_2}$ t_i' $[_{C_e}$ hat_j Ø] $[_{TP}$ sie t_i getroffen t_j]]]]				

*T_6: *Wh-movement from V/2 clauses into 'dass' clauses:* (3-d)

Input: $[_{C_{dw}}$ $dass_{[*wh*]}$], $[_{TP}$ du meinst $[_{CP_2}$ wen_i $[_{C_e}$ hat_j Ø] $[_{TP}$ sie t_i getroffen t_j]]]	FC	OP	SCE	LR
O_{41}: $[_{CP_1}$ $[_{C_{dw}}$ dass] $[_{TP}$ du meinst $[_{CP_2}$ wen_i $[_{C_e}$ hat_j Ø] $[_{TP}$ sie t_i getroffen t_j]]]]	*!		**	
O_{42}: $[_{CP_1}$ wen_i $[_{C_{dw}}$ dass] $[_{TP}$ du meinst $[_{CP_2}$ t_i' $[_{C_e}$ hat_j Ø] $[_{TP}$ sie t_i getroffen t_j]]]]			*!	
☞O_{43}: $[_{CP_1}$ wen_i $[_{C_{dw}}$ $meinst_l$ (dass)] $[_{TP}$ du t_l $[_{CP_2}$ t_i' $[_{C_e}$ hat_j Ø] $[_{TP}$ sie t_i getroffen t_j]]]]				*

In a nutshell, then, the present analysis of the pattern in (3) is this: Given an optimal CP_2, SCE demands that the edge of CP_1 has the same shape. This requirement can be met without problems in (3-a) and (3-c), where C_1 and C_2 are uniformly marked *d* (*dass*) or *e* (V/2). However, in (3-b) and (3-d), C_1 and C_2 differ with respect to *d*/*e* marking. This means that SCE can only be satisfied by violating some other constraint. In (3-b), this other constraint is

‡ C_e and C_{dw} in O_{43} of T_6 have identical shapes, as branching Cs. Note that the neutralization effect is not complete since O_{44} of T_5 has a Ø where O_{43} of T_6 has a *dass*. Hence, we would have to assume obligatory *dass* deletion at PF if O_{43} of T_6 could be (part of) a well-formed derivation – which, however, it can't be: O_{43} of T_6 can only be an intermediate optimal output (C_{dw} is always embedded and cannot be the head of a root clause); and there is a general prohibition against embedded *wh*-V/2 constructions in German (see Haider (1984)):

(i) a. *Sie sagt $[_{CP}$ wen_i meinst du t_i]
  she says who mean you

  b. Sie sagt $[_{CP}$ wen_i du t_i meinst]
  she says who you mean

Accordingly, merging the optimal CP output of T_6 (or T_5, for that matter) with V invariably results in ungrammaticality. Thus, independently of present considerations, there must be a high-ranked constraint $\Re$ against merging V and a CP with $XP_{[wh]}$-V/2 at its edge. Ineffability can then be derived in this context under a ranking $\Re \gg$ EOC, where EOC is the *Empty Output Condition* that blocks the empty output Ø (the null parse). Ø is always present in competitions; its optimality signals a crash of the derivation (see Heck and Müller (2000a)).

FC; in (3-d), it is LR. Consequently, the ranking FC $\gg$ SCE $\gg$ LR correctly predicts that SCE cannot stop feature-driven V/2 from applying in (3-b) (T_3), and that SCE forces repair-driven V/2 in (3-d) (T_6).

Needless to say, there are several further questions that will have to be addressed before the analysis can count as successful, and it will have to be extended in various ways.§ Still, I would like to contend that the gist of the analysis in T_1–T_6 can be maintained in a more comprehensive approach.

5. Argument

It remains to be shown that a global optimization approach would, ceteris paribus, fail in an analysis of the pattern in (3). This is straightforward: Under a global approach, we would wrongly expect SCE to require identity of the shape of clause edges much more generally, and could not account for the asymmetry observed in (3). In particular, (3-b) should be excluded in the same way as (3-d): CP_1 in O_{24} of T_3 violates SCE once; its predecessor CP_2 in O_2 of T_1 violates LR once. However, if the two CPs are optimized in parallel, the optimal output would combine CP_1 in O_{24} of T_3 and CP_2 in O_3 of T_1 (which is locally suboptimal because of a fatal LR violation due to locally unforced V/2). This would incur two violations of LR, but *no violation of SCE*; see T_7, where the wrong winner (O_{34}, based on O_3) is marked ★, and well-formed O_{24} is blocked because of a fatal SCE violation. More generally, the global optimization approach predicts that an output at the right end of a non-arrowed line at a level like Σ_2 in figure 1 can be further used, and may ultimately lead to an output at a later level like Σ_3 that has a better constraint profile than the corresponding output at the right end of an arrowed line. This prediction is not borne out, though; hence, we have an argument against global optimization.

*T_7: Global optimization: *Wh-movement from 'dass' clauses into V/2 clauses:* (3-b)

Input: [$_{C_d}$ dass], [$_{TP}$ sie wen getroffen hat] [$_{C_{ew}}$ Ø$_{[*wh*],[*fin*]}$], [$_{TP}$ du meinst]	FC	OP	SCE	LR
O_{24}: [$_{CP_1}$ wen$_i$ [$_{C_{ew}}$ meinst$_l$ Ø] [$_{TP}$ du t$_l$ [$_{CP_2}$ t$'_i$ [$_C$ dass] [$_{TP}$ sie t$_i$ getroffen hat]]]]			*!	*
★O_{34}: [$_{CP_1}$ wen$_i$ [$_{C_{ew}}$ meinst$_l$ Ø] [$_{TP}$ du t$_l$ [$_{CP_2}$ t$'_i$ [$_C$ hat$_j$ (dass)] [$_{TP}$ sie t$_i$ getroffen t$_j$]]]]				**

6. Outlook

I have argued that a local approach to optimization in syntax is conceptually superior to a global approach because it reduces complexity; and I have shown that it also proves

§ To name just one relevant question: Why does SCE not force XP movement to SpecC in a matrix clause in the presence of ([*xp*]-driven) XP movement to SpecC in an embedded clause? In this context, there is no asymmetry between embedded *dass* clauses (as in (i-a)) and V/2 clauses (as in (i-b)); in particular, there is no repair-driven movement of both *er* and *sagte* to the edge of CP_1 in (i-b).

(i) a. Ich denke [$_{CP_1}$ er [$_{C_e}$ sagte] [$_{CP_2}$ [$_{C_d}$ dass] sie schlafen möchte]]
   I think he said that she sleep wants to
  b. Ich denke [$_{CP_1}$ [$_{C_d}$ dass] er sagte [$_{CP_2}$ sie [$_{C_e}$ möchte] schlafen]]
   I think that he said she wants to sleep

A simple solution would be to postulate a constraint $\Im$ ($\Im \gg$ SCE) that permits movement of a non-operator to the edge of C only if C is marked [*xp*]. On this view, movement theory is designed in such a way that only those items can move successive-cyclically that do in fact need to move in this manner, viz., operators.

empirically superior in the domain of successive-cyclic movement from *dass* vs. V/2 CPs in German, where it solves a recalcitrant problem via a simple Shape Conservation constraint.

The question arises of whether the local approach to optimization in syntax can be maintained in its strictest form (without adding limited look-ahead or backtracking capacity) in the light of other constructions that involve long-distance dependencies and thereby initially seem to support to a global approach. Phenomena that are relevant in this context include non-local reflexivization and resumptive pronoun strategies. Such non-local binding phenomena will have to be handled in a local approach by systematically decomposing non-local relations into a series of local feature passing operations (as proposed in Gazdar et al. (1985)), such that relevant information is accessible in each local optimization procedure. At the moment, I take it to be an open question whether this enterprise will ultimately be successful; however, preliminary results (see, e.g., Fischer (2003) on reflexives) suggest that such apparently non-local phenomena can indeed fruitfully be addressed in a local approach to optimization.

7. Bibliography

Ackema, Peter and Ad Neeleman (1998): Optimal Questions, *Natural Language and Linguistic Theory* 16, 443–490.

Aissen, Judith (2002): Bidirectional Optimization and the Problem of Recoverability in Head Marking Languages. Ms., University of California, Santa Cruz.

Archangeli, Diana and Terence Langendoen (1997): Afterword. In: D. Archangeli and T. Langendoen, eds., *Optimality Theory. An Overview*. Blackwell, Oxford, pp. 200–215.

Barbosa, Pilar, Danny Fox, Paul Hagstrom, Martha McGinnis and David Pesetsky, eds. (1998): *Is the Best Good Enough?*. MIT Press and MITWPL, Cambridge, Mass.

Blutner, Reinhard (2000): Some Aspects of Optimality in Natural Language Interpretation, *Journal of Semantics* 17, 189–216.

Bresnan, Joan (2001): The Emergence of the Unmarked Pronoun. In: G. Legendre, J. Grimshaw and S. Vikner, eds., *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Mass., pp. 113–142.

Broekhuis, Hans (2000): Against Feature Strength: The Case of Scandinavian Object Shift, *Natural Language and Linguistic Theory* 18, 673–721.

Chomsky, Noam (1973): Conditions on Transformations. In: S. Anderson and P. Kiparsky, eds., *A Festschrift for Morris Halle*. Academic Press, New York, pp. 232–286.

Chomsky, Noam (2000): Minimalist Inquiries: The Framework. In: R. Martin, D. Michaels and J. Uriagereka, eds., *Step by Step*. MIT Press, Cambridge, Mass., pp. 89–155.

Chomsky, Noam (2001): Derivation by Phase. In: M. Kenstowicz, ed., *Ken Hale. A Life in Language*. MIT Press, Cambridge, Mass., pp. 1–52.

Fanselow, Gisbert and Damir Ćavar (2001): Remarks on the Economy of Pronunciation. In: G. Müller and W. Sternefeld, eds., *Competition in Syntax*. Mouton de Gruyter, Berlin, pp. 107–150.

Fischer, Silke (2002): Reanalyzing Reconstruction Effects: An Optimality-Theoretic Account of the Relation between Pronouns and R-Expressions. In: M. van Koppen, E. Thrift, E. J. van der Torre and M. Zimmerman, eds., *Proceedings of ConSole 9*. HIL, Leiden, pp. 69–81.

Fischer, Silke (2003): Binding in a Derivational Approach. Ms., Universität Stuttgart. (Chapter 5 of forthcoming dissertation.).

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum and Ivan Sag (1985): *Generalized Phrase Structure Grammar*. Blackwell, Oxford.

Grimshaw, Jane (1997): Projection, Heads, and Optimality, *Linguistic Inquiry* 28, 373–422.

Haider, Hubert (1984): Topic, Focus, and V-Second, *Groninger Arbeiten zur Germanistischen Linguistik* 25, 72–120.

Haider, Hubert (1993): ECP-Etüden: Anmerkungen zur Extraktion aus eingebetteten Verb-Zweit-Sätzen, *Linguistische Berichte* 145, 185–203.

Heck, Fabian (2001a): Pied Piping Without Feature Percolation. Ms., Universität Stuttgart.

Heck, Fabian (2001b): Quantifier Scope in German and Cyclic Optimization. In: G. Müller and W. Sternefeld, eds., *Competition in Syntax*. Mouton/de Gruyter, Berlin, pp. 175–209.

Heck, Fabian and Gereon Müller (2000a): Repair-Driven Movement and the Local Optimization of Derivations. Ms., Universität Stuttgart and IDS Mannheim.

Heck, Fabian and Gereon Müller (2000b): Successive Cyclicity, Long-Distance Superiority, and Local Optimization. In: R. Billerey and B. D. Lillehaugen, eds., *Proceedings of WCCFL*. Vol. 19, Cascadilla Press, Somerville, MA, pp. 218–231.

Hendriks, Petra and Helen de Hoop (2001): Optimality Theoretic Semantics, *Linguistics and Philosophy* 24, 1–32.

Hermans, Ben and Mark van Oostendorp, eds. (2000): *The Derivational Residue in Phonological Optimality Theory*. Benjamins, Amsterdam.

Jäger, Gerhard (2002): Learning Constraint Sub-Hierarchies. The Bidirectional Gradual Learning Algorithm. Ms., ZAS Berlin/Universität Potsdam.

Jäger, Gerhard and Reinhard Blutner (2000): Against Lexical Decomposition in Syntax. In: A. Wyner, ed., *Proceedings of IATL*. Vol. 15, University of Haifa, pp. 113–137.

Lee, Hanjung (2001): Optimization in Argument Expression and Interpretation: A Unified Approach. PhD thesis, Stanford University.

Legendre, Géraldine, Jane Grimshaw and Sten Vikner, eds. (1998): *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Mass.

Legendre, Géraldine, Paul Smolensky and Colin Wilson (1998): When is Less More? Faithfulness and Minimal Links in Wh-Chains. In: P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky, eds., *Is the Best Good Enough?*. MIT Press and MITWPL, Cambridge, Mass., pp. 249–289.

McCarthy, John (2000): Harmonic Serialism and Parallelism. In: M. Hirotani, A. Coetzee, N. Hall and J.-Y. Kim, eds., *Proceedings of NELS 30*. GLSA, Amherst, Mass., pp. 501–524.

Müller, Gereon (1989): Barrieren und Inkorporation. Master's thesis, Universität Konstanz.

Müller, Gereon (2000): Shape Conservation and Remnant Movement. In: M. Hirotani, A. Coetzee, N. Hall and J.-Y. Kim, eds., *Proceedings of NELS 30*. GLSA, Amherst, Mass., pp. 525–539.

Müller, Gereon (2001): Order Preservation, Parallel Movement, and the Emergence of the Unmarked. In: G. Legendre, J. Grimshaw and S. Vikner, eds., *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Mass., pp. 279–313.

Müller, Gereon (2002): Harmonic Alignment and the Hierarchy of Pronouns in German. In: H. Simon and H. Wiese, eds., *Pronouns: Grammar and Representation*. Benjamins, Amsterdam, pp. 205–232.

Müller, Gereon and Wolfgang Sternefeld (1993): Improper Movement and Unambiguous Binding, *Linguistic Inquiry* 24, 461–507.

Perlmutter, David and Scott Soames (1979): *Syntactic Argumentation and the Structure of English*. The University of California Press, Berkeley.

Pesetsky, David (1998): Some Optimality Principles of Sentence Pronunciation. In: P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis and D. Pesetsky, eds., *Is the Best Good Enough?*. MIT Press and MITWPL, Cambridge, Mass., pp. 337–383.

Prince, Alan and Paul Smolensky (1993): Optimality Theory. Constraint Interaction in Generative Grammar. Book ms., Rutgers University.

Pullum, Geoffrey (1979): *Rule Interaction and the Organization of a Grammar*. Garland, New York.

Reis, Marga (1985): Satzeinleitende Strukturen im Deutschen. In: W. Abraham, ed., *Erklärende Syntax des Deutschen*. Narr, Tübingen, pp. 271–311.

Reis, Marga (1996): Extractions from Verb-Second Clauses in German?. In: U. Lutz and J. Pafel, eds., *On Extraction and Extraposition in German*. Benjamins, Amsterdam, pp. 45–88.

Rubach, Jerzy (2000): Glide and Glottal Stop Insertion in Slavic Languages: A DOT Analysis, *Linguistic Inquiry* 31, 271–317.

Sells, Peter, ed. (2001): *Formal and Empirical Issues in Optimality Theoretic Syntax*. CSLI, Stanford.

Staudacher, Peter (1990): Long Movement from Verb-Second-Complements in German. In: G. Grewendorf and W. Sternefeld, eds., *Scrambling and Barriers*. Benjamins, Amsterdam, pp. 319–339.

Sternefeld, Wolfgang (1989): V-Movement, Extraction from V/2 Clauses, and the ECP, *Working Papers in Scandinavian Syntax* 44, 119–140.

Sternefeld, Wolfgang (1992): Transformationstypologie und strukturelle Hierarchie. Ms., Universität Tübingen.

Sternefeld, Wolfgang (2003): Syntax. Eine merkmalsbasierte Analyse. Book ms., Universität Tübingen.

Tappe, Thilo (1981): Wer glaubst du hat recht?. In: M. Kohrt and J. Lenerz, eds., *Sprache: Formen und Strukturen*. Niemeyer, Tübingen, pp. 203–212.

Vogel, Ralf (2002): Feedback Optimisation. Economy and Markedness in OT Syntax, Pt. I. Ms., Universität Potsdam.

Williams, Edwin (1974): Rule Ordering in Syntax. PhD thesis, MIT, Cambridge, Mass.

Williams, Edwin (2003): *Representation Theory*. MIT Press, Cambridge, Mass.

Wilson, Colin (2001): Bidirectional Optimization and the Theory of Anaphora. In: G. Legendre, J. Grimshaw and S. Vikner, eds., *Optimality-Theoretic Syntax*. MIT Press, Cambridge, Mass., pp. 465–507.