

Correspondence in OT Syntax and Minimal Link Effects*

Ralf Vogel

Abstract

The aim of this paper is the exploration of an optimality theoretic architecture for syntax that is guided by the concept of *correspondence*: syntax is understood as the mechanism of “translating” underlying representations into a surface form. In minimalism, this surface form is called “Phonological Form” (PF). Both semantic and abstract syntactic information are reflected by the surface form. The empirical domain where this architecture is tested are minimal link effects, especially in the case of *wh*-movement.

The OT constraints require the surface form to reflect the underlying semantic and syntactic representations as maximally as possible. The means by which underlying relations and properties are encoded are precedence, adjacency, surface morphology and prosodic structure. Information that is not encoded in one of these ways remains unexpressed, and gets lost unless it is recoverable via the context. Different kinds of information are often expressed by the same means. The resulting conflicts are resolved by the relative ranking of the relevant correspondence constraints.

The minimal link condition (cf. Chomsky 1995, Rizzi 1990) as given in (1) expresses a locality restriction on syntactic movement: Movement of α to a target K is blocked by β , if β is closer to K and could enter the same checking relation.

- (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 .
(Chomsky 1995, 311)

The restriction described by the MLC has been explained in two ways. The most common explanation is in terms of *economy of movement*: β blocks α because movement of β would require a *shorter* movement step. Economy of movement is a core principle of grammar in minimalist syntax (Chomsky, 1995). An optimality theoretic implementation of this idea has been developed by Legendre et al. (1998) in the form of the constraint BAR :

- (2) BAR : A chain link may not cross a barrier.
(Legendre et al. 1998, 261, see also Hale and Legendre, this volume)

Conjoined versions of BAR like BAR^2 (“A chain link may not cross two barriers.”), BAR^3 etc. make up the “ MINLINK power hierarchy”. The more barriers are crossed, the more violations of BAR constraints are incurred by a movement step. Candidates with fewer BAR violations block others with more violations.

The second strategy of explanation for the MLC relies on the fact that movement of α across β reverses the relative order of these two elements. Within OT, Müller (2001) presented an analysis of various syntactic

* This paper has grown over a longer period of time. I want to thank everyone who helped me clarify my thoughts on these issues: Gereon Müller, Sten Vikner, Fabian Heck, Tanja Schmid, Silke Fischer, Jonas Kuhn, Gunnar Hrafnbjargarson, Gisbert Fanselow, Arthur Stepanov, Joanna Blaszczak, Susann Fischer, Florian Schäfer, Eva Engels, Andreas Haida, Hanjung Lee, Yukiko Morimoto, Gerhard Jäger, Doug Saddy, Peter Staudacher, Hubert Haider, Hans-Martin Gärtner, Cornelia Endriss, Jane Grimshaw, Judith Aissen, Juan Uriagareka, Vieri Samek-Lodovici, John Hale, Géraldine Legendre, three anonymous reviewers, and the audiences of several presentations in Potsdam in November 2001 (WOTS 5), July 2002 (minimal link workshop), and in Cologne (January 2003). The copyright for all errors and shortcomings contained in this paper is mine, however. The work on this paper has been supported by the Deutsche Forschungsgemeinschaft, research group “Conflicting Rules in Language and Cognition”, FOR-375/2-A3.

movement phenomena that seem to be governed by the force to keep the underlying relative order of elements in a surface form. Müller calls the constraint ‘Parallel Movement’:

- (3) ‘Parallel Movement’ (PAR-MOVE):
 If α c-commands β at level L_n , then α c-commands β at level L_{n+1}
 (where β, α are arguments).
 (Müller 2001, 279)

Williams’s (2003) ‘Representation Theory’ is largely built on the principle of *shape conservation*. His proposal is more radical than Müller’s in that shape conservation is conceived as a *replacement* for derivational economy.

Under the term ‘*faithfulness*’, structure preservation already plays a central role in OT. *Correspondence Theory*, as developed by McCarthy and Prince (1995), is an extension and systematisation of the standard input-output faithfulness system. Relations between representations, elements within representations and their properties are evaluated by a set of correspondence constraints, demanding, among other things, the existence of a correspondent, the conservation of the relative order of elements, one-to-one mappings, feature identity of corresponding elements etc.

If it is possible to reconstruct the MLC as shape conservation, then this is a natural way of dealing with it within OT, because one would only use tools which are already there, while derivational economy needs to be added in the form of constraints like BAR, or Grimshaw’s (1997) STAY (‘No movement’), or by other means.

An OT account of the MLC in terms of derivational economy still needs correspondence. Let me demonstrate this with the example of superiority in English:

- (4) **What did who say?*

This is one of the standard cases for which the MLC has been used as explanation. Movement of the object to the clause-initial position [Spec,CP] involves the crossing of more barriers than movement of the subject. Consider a syntactic representation where case is assigned in the specifier of AGR-phrases, as it is usual in minimalist syntax. The table in (5) displays the BAR violations of possible OT candidates for (4):

(5)

Qy.Qx.say(x,y)	BAR
c1: [CP W _O ... [AGRSP W _S [AGROP t _O [VP t _S t _O]]]]	*****!
☞ c2: [CP W _S ... [AGRSP t _S [AGROP W _O [VP t _S t _O]]]]	****
☞ c3: [CP W _O ... [AGROP t _O [AGRSP W _S [VP t _S t _O]]]]	****

Candidate c1 represents the ungrammatical (4). The object *wh*-phrase crosses one barrier, VP, when moving to AGROP, and two further ones, AGROP and AGRSP, when moving to CP. The subject movement crosses VP and AGROP.¹ We have five violations of BAR. The blocking candidate c2 has one violation less because the final subject movement to CP crosses only one barrier. But what about candidate c3? Here, we reversed the relative order of AGROP and AGRSP, and we see that we have only four violations now. This candidate has the ungrammatical surface order in (4) and

¹Legendre et al. (1998) use Chomsky’s (1986) definition of barrier, where IP is a barrier by inheritance from VP. This is here carried over to all functional projections above VP.

therefore must not be a winner.

The problem that is raised here for the OT implementation of the derivational MLC is: If β blocks movement of α , why does β precede α in the first place? Could there not be a candidate that has the reverse underlying order? Furthermore, if movement to [Spec,CP] is blocked, why not *insert* α directly into that position without movement? In OT, with its relatively unconstrained candidate generator, these are real options that must be excluded explicitly. This is usually achieved by stipulating that AGRSP must embed AGROP universally, that all NPs have to be inserted into their “theta positions” within VP etc. – inviolable constraints that are assumed to be part of the candidate generator, GEN.

A theory that is built on shape conservation has a straightforward explanation for the restrictions just discussed. In LFG-OT, they follow from f-structure/c-structure correspondence (see Kuhn 2001). In the system of Williams (2003), case structure represents theta structure, and the parallel relative order of corresponding elements follows from shape conservation. This is certainly the most plausible explanation for the parallelism of case and argument hierarchy.

A correspondence theoretic OT approach can use ordinary violable OT constraints here which are not part of GEN. This is the second conceptual advantage of such a model. GEN should be as unconstrained and do as little explanatory work as possible: the assumption of inviolable constraints as such weakens an OT account. Limitations of space do not allow an exhaustive discussion of all details of an OT syntax model based on correspondence. In what follows, I will concentrate on those aspects that are relevant for the discussion of minimal link effects, first of all superiority.

The paper is organised as follows: In section 1 the architecture of the proposed OT syntax model is laid out; section 2 demonstrates in a first application how Greenberg’s (1963) first universal can be derived. Section 3 analyses English topicalisation and *wh*-movement in English, in section 4 this is also done for German. The discussion in section 5 focuses on multiple questions. Section 6 focuses on word order freezing, introduces recoverability as a central criterion for grammaticality, and shows how this can be implemented in a bidirectional model of the grammar.

1 Correspondence

1.1 A brief sketch of Optimality Theory

Optimality Theoretic models consist of five components:

- An *input* representation I_n
- A set of representations of *output candidates* O .
- A *generation* function $GEN(I_n, O)$ that generates O on the basis of I_n .
- A constraint hierarchy CON .
- An *evaluation* function $EVAL(CON, O, O_n)$ that selects the optimal output O_n from O on the basis of CON .

This model can be applied to very different tasks. The input and output representations often vary with the problem that is targeted. In OT syntax, the input is often considered a more or less complex semantic representation, and the output a syntactic representation. Other approaches use an abstract syntactic representation for the input, and a phonological representation as the output. Input and output representations are as unrestricted as possible. Constraints should be part of CON .

The constraints are violable and hierarchically ordered. Low ranked constraints can never override the effects of higher ranked constraints, no matter how often they are violated. There are two constraint types: *Markedness* constraints evaluate features of candidates, while *faithfulness* constraints evaluate how similar an output candidate is to the input. This latter constraint type is particularly important for our discussion. Output candidates could be more complex, for instance, a pair of representations. In that case, correspondence of two parts of an output candidate is also an issue. A generalised theory of correspondence that includes input-output faithfulness as a special case has been presented by McCarthy and Prince (1995).

1.2 Correspondence-based OT syntax

Müller's (2001) PAR-MOV requires correspondence between different levels of syntactic representations. He uses the standard model from Chomsky's (1981) Government and Binding Theory with D-structure, S-structure and LF as levels. An alternative that is based on more recent minimalist syntax has been proposed by Heck and Müller (2000) in their system of serial optimisation.

A second kind of correspondence is also conceivable, namely, one that relates different kinds of representations. Standard OT syntax, as introduced by the work of Grimshaw (1997) and others², assumes the input to be a semantic representation containing first of all the argument structure, but also scope representations for operators and quantifiers. Most researchers also include information structural specifications in the input, as has been suggested by work dealing with the relation between syntax and information structure.³

The candidates are syntactic representations. Faithfulness constraints in this case evaluate the correspondence of semantic and syntactic representations. Other examples for correspondence between different types of representations are the papers by Pesetsky (1997,1998), where the syntactic structure stands in correspondence with the phonological output representation, PF. Legendre (2000) uses constraints on syntax-PF correspondence in her essentially PF-based analysis of the behaviour of Bulgarian clitics. These are examples for a usage of the term "correspondence" as it is prominent in the work of Jackendoff (1990). The role of syntax, according to Jackendoff, is to mediate between meaning and sound, semantics and phonology, by virtue of so-called "correspondence rules".

I want to explore an architecture for OT syntax which is based on Jackendoff's notion of correspondence, however, I also assume that there must be a direct correspondence relation between conceptual and phonological structures, as has been proposed, for instance, in work on the relation between prosodic and information structure. I will use the labels M ("meaning", semantics), S (abstract syntax) and P (phonological representation or surface syntax).

The mapping relations that hold between these representations have the character of a *translation*. For instance, scope is often translated into c-command and precedence, the same holds for other instances of prominence, like being the higher argument or being more salient in the discourse. Sisterhood and adjacency is a similar pair. Truckenbrodt (1999) discusses the relation between syntactic and phonological phrases.

²See the collection by Legendre et al. (2001) for a representative overview. The state of the art in LFG-oriented OT syntax is documented in (Bresnan, 2000) and the collection edited by Peter Sells (2001).

³(Choi, 1996) and (Büring, 2001) are two examples in case.

Following more recent work in minimalist syntax, I assume that the abstract syntactic representation does not represent linear order itself. This is only represented at P. Kayne (1994) postulates the *Linear Correspondence Axiom* which expresses, as a mapping principle, that asymmetric c-command in a syntactic representation translates into precedence. In the system explored here, this is a good candidate for a correspondence constraint, but, as an OT constraint, it would be violable.⁴

I assume that S is an X-bar structure, where phrases have a top phrasal node, at most one specifier, at most one X-bar node dominating a head and at most one complement. Movement operations are restricted to cyclic substitution. All (abstract) syntactic movement is therefore structure-building.

Adjunction is treated as an instance of *linear reordering* in the mapping from S to P. Consequently, true syntactic adjuncts, like adverbial modifiers, are generated together with matrix trees, but not connected to them in S. So there is no abstract syntactic operation of adjunction, neither of XPs nor of heads. S therefore is not necessarily a unique syntactic tree, but could as well be a ‘forest’, a set containing the matrix tree and all its adjuncts as elements.⁵

As long as there is movement, there is also the notion of movement chain. I assume that only the head of a movement chain is relevant for the mapping from S to P. This might help in reducing movement and assumptions about it to a minimum. But this restriction might as well follow from a *recoverability* condition: how can a movement step be made visible, if the trace is spelled out instead of the movement chain’s head?

S and P could be seen as two representations that both partially represent syntactic properties of a clause. While S represents constituency, abstract features and further abstract syntactic relations (like binding, case assignment, and others, insofar as they are assumed to be syntactic), linear order is represented at P, together with other surface aspects, like morphology and prosodic phrasing.

⁴A brief remark about frameworks: S, the abstract syntactic representation, is encoded within X-bar theory in this paper. But this is not crucial for the proposed analysis. S could as well be an HPSG feature structure, or a combination of f-structure and c-structure in LFG (without linearisation). The choice of representational alphabet is not of the same importance within OT syntax than in non-OT syntax, because all explanation lies or should lie in the constraints and their interaction. The constraints used in this paper deal with the correspondence between representations and are therefore formulated relative to the representational language that is used, but I expect that the constraints can be recoded to work with HPSG, LFG, or any other syntactic framework.

⁵In a syntax class in winter 2001/2002, University of Potsdam, Juan Uriagareka gave a very similar formulation of this idea in correlation with his conception of ‘multiple spell-out’ (Uriagareka, 1999): An adjunct is generated in parallel to the matrix tree, and adjoined, i.e. adjacent to that part of the tree that it is spelled out together with. A similar proposal within minimalism is Stepanov’s (2001) theory of *Late Adjunction* that treats adjunction as post-cyclic operation. Chomsky (2001, 16) seems to have something similar in mind when he states that ‘given the basic properties of adjunction, we might intuitively think of α as attached to β on a separate plane, with β retaining all its properties on the ‘primary plane’, the simple structure. ...’. The notion of ‘separate plane’ remains unclear. My proposal suggests that it is P.

A question that immediately arises, and which has also been brought up by one anonymous reviewer, is how ditransitive verbs are represented in this proposal. One possibility is Larson’s (1988) solution of a ‘VP-Shell’, where the necessary positions for the NPs are provided by several stacked projections of the same verb. This might be the correct proposal for English. However, it has also been argued that arguments can be syntactic adjuncts. Grimshaw (1990) calls the English ‘by’-phrase an ‘argument adjunct’. Likewise, Vogel and Steinbach (1998) present rich evidence that German dative objects should best be treated as adjuncts syntactically. Both of these options are possible.

I will introduce *recoverability* below as another central aspect of the correspondence-based OT syntax model. M and S are *encoded* into P and must be *completely recoverable* from P. This defines grammaticality. Four constraint families form the base of this system. These are constraints on the correspondence of M and S ($M \Rightarrow S$), M and P ($M \Rightarrow P$), and S and P ($S \Rightarrow P$). The fourth family, $S_i \Rightarrow S_o$, will become necessary with the inclusion of S in both input and output (see section 5 below).

2 Deriving Greenberg's First Universal

As a first demonstration, I want to show how such a system is able to derive Greenberg's (1963) first universal:

- (6) *Universal I*. In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object.

This universal proposes three basic word order patterns, VSO, SVO and SOV. In the meantime, it has become clear that VOS is also an option. Malagasy (Rackowski and Travis, 2000) and Tzotzil (Aissen, 1987; , 1992; , 1996) are examples.⁶ It seems that in some Mayan languages clauses of the form 'V NP NP' are ambiguous for VSO and VOS order, if no other features of the NPs disambiguate the structure. The examples in 7 are from K'ichee' (Mondloch, 1978):⁷

- (7) VSO/VOS order in K'ichee' (Mayan):
 a. *X-uu-kuna-j rii achih rii ixoq.*
 CP-E3-cure-ACT DET man DET woman
 'The man cured the woman' or 'The woman cured the man'.
 b. *X-ee-ki-kuna-j rii achijaab' rii ixoqiib'.*
 CP-A3pl-E3pl-cure-ACT DET men DET women
 'The men cured the women' or 'the women cured the men'.
 (CP = completive aspect; E = ergative; A = absolutive; ACT = suffix for active transitive verbs; DET = determiner)

In the following, I assume that VOS is an option and should be a possible winner under some ranking. The orders that have to be excluded as default orders are OVS and OSV. The constraints in 9 evaluate how a VP of the form 8 is linearised. They are introduced in an informal way and will be given more precise definitions in section 4.⁸

- (8) [VP Subject [_v V⁰ Object]]
- (9) $S \Rightarrow P(\text{sh})$: A specifier precedes its head.
 $S \Rightarrow P(\text{ch})$: A complement precedes its head.
 $S \Rightarrow P(\text{hc})$: A head precedes its complement.

⁶For a recent overview and an alternative treatment within OT, see Zepter (2003).

⁷I am very thankful to Judith Aissen for providing me with information about these facts, and for discussing this and other aspects of the approach presented here. See also Zepter (2003) for an alternative OT treatment of these issues.

⁸Some of these constraints have predecessors in earlier work. Grimshaw's (1997) 'HeadLeft' and 'HeadRight' closely resemble $S \Rightarrow P(\text{hc})$ and $S \Rightarrow P(\text{ch})$. In general, these constraints express common generalisations about unmarked orders.

S⇒P(sa): Sisters are adjacent.

S⇒P(NP): Asymmetric c-command among NPs translates into precedence.

M⇒P(fa): A functor precedes all its arguments.

The six logically possible linearisations of subject, object and verb constitute the candidate set. Four of the six candidates are possible under some ranking.⁹ Two orders are correctly excluded, those with the object in first position. This is not to say that such orders are impossible, they are only impossible as default or unmarked orders. Table 10 displays the violations of the constraints by the six candidates and shows how this result is achieved.

(10)

[VP S [v' V ⁰ O]]	S⇒P (sh)	S⇒P (ch)	S⇒P (hc)	S⇒P (sa)	S⇒P (NP)	M⇒P (fa)
VSO	*	*		*		
VOS	*	*			*	
SVO		*				*
SOV			*			*
OSV			*	*!	*!	*
OVS	*!		*		*!	*

The orders OSV and OVS violate the same constraints as the SOV order, and further constraints in addition which have been marked with “!”. The candidates are ‘harmonically bounded’: they would lose against SOV under any ranking. A constraint that is violated by both is S⇒P(NP). This is the constraint that formulates the core of Greenberg’s first universal, namely, that subjects precede objects, i.e. that asymmetric c-command between elements of the syntactic category NP translates into precedence. The only candidate that violates this constraint and is not harmonically bounded, is VOS. The reason is that VSO, the candidate with the most similar violation profile, cannot fulfil sister adjacency: head initial orders cannot simultaneously fulfil S⇒P(NP) and S⇒P(sa). The winning rankings for the four possible winners are summarised in 11 – only the crucial rankings are indicated:

- (11) SVO: {S⇒P(hc), S⇒P(sa), S⇒P(NP), S⇒P(sh)} >> {S⇒P(ch), M⇒P(fa)}
 SOV: {S⇒P(ch), S⇒P(sa), S⇒P(NP), S⇒P(sh)} >> {S⇒P(hc), M⇒P(fa)}
 VSO: {S⇒P(hc), M⇒P(fa), S⇒P(NP), S⇒P(sh)} >> {S⇒P(ch), S⇒P(sa)}
 VOS: {S⇒P(hc), M⇒P(fa), S⇒P(sa)} >> {S⇒P(ch), S⇒P(NP), S⇒P(sh)}

A language like K’ichee’ would have a constraint ranking where S⇒P(NP) and S⇒P(sa) are *tied*, i.e., are of equal rank:¹⁰

⁹The typological predictions of this system of constraints have been calculated with the help of ‘OTSoft’ (Hayes et al., 2002).

¹⁰Several definitions for constraint ties have been proposed. The least problematic one, which is sufficient here, is the one that interprets a tie as an abbreviation for the existence of two co-grammars which only differ in the relative ranking of the two tied constraints. The advantage of this definition is that it does not introduce a new kind of constraint interaction. The tie is only a notational convention. McCarthy (2002, 227) traces this definition back to Paul Kiparsky, who motivated it with considerations about language change.

- (12) VSO/VOS (e.g., K'ichee'):
 $\{S \Rightarrow P(\text{hc}), M \Rightarrow P(\text{fa})\} \gg S \Rightarrow P(\text{sa}) \circ S \Rightarrow P(\text{NP}) \gg S \Rightarrow P(\text{ch})$

Head-initial orders cannot fulfil both $S \Rightarrow P(\text{sa})$ and $S \Rightarrow P(\text{NP})$ simultaneously, contrary to SVO and SOV. A tie between $S \Rightarrow P(\text{sa})$ and $M \Rightarrow P(\text{fa})$ in combination with ' $S \Rightarrow P(\text{hc}) \gg S \Rightarrow P(\text{ch})$ ' would yield a VSO/SVO language. Spanish has been classified as SVO and VSO by different authors.¹¹ Perhaps, both orders are equally unmarked. A tie between $S \Rightarrow P(\text{sa})$ and $M \Rightarrow P(\text{fa})$ in combination with ' $S \Rightarrow P(\text{hc}) \gg S \Rightarrow P(\text{ch})$ ' would yield such a VSO/SVO language.

The following sections will explore how the system of constraints described in this section implements the minimal link condition. The discussion will focus on A-bar movement, in particular, topicalisation and *wh*-movement. I will compare the accounts for English and German.

3 Topicalisation and *wh*-movement in English

In the English sentences in (13), the order of the NPs is determined by argument structure, operator scope and information structure, respectively.¹²

- (13) a. *John wrote this book.* order follows argument structure
 b. *What did John write?* order follows *wh*-scope marking
 c. *The red book, John wrote.* order follows information structure

In both (13b) and (13c), the order of the arguments as determined by argument structure is not preserved. These examples could be seen as simple cases of 'minimal link violations', under a version of the minimal link condition that is equal to the requirement to preserve the underlying argument order. Many cases of MLC violations discussed in the literature are of this kind.

What is particularly interesting is that (13b) and (13c) have two different syntactic structures: while in (13b) '*what*' is moved into the specifier of an additional projection headed by '*do*', '*The red book*' in (13c) is left dislocated, presumably adjoined to the root node. The two different structural solutions can be interpreted as resulting from different priorities among correspondence constraints. Scope marking is reflected at both S and P, while information structural prominence is only reflected at P. I assume the following constraints on the correspondence between M and S:¹³

- (14) Constraints on $M \Rightarrow S$ mapping:
 (identical indices indicate correspondence of elements, e.g., m_1 corresponds to s_1)
 a. $M \Rightarrow S(\text{Arg})$: If an argument m_1 is higher than another argument m_2 at M, then s_1 asymmetrically c-commands s_2 at S.

¹¹See Costa (2001) for some discussion and the references cited there.

¹²Prince (1999) shows that calling the fronting operation in (13c) 'topicalisation' is misleading in that the fronted element usually does not function as a discourse topic. The operation is possible with focus bearing phrases in particular discourse configurations. Without going into detail, I will refer to this particular property as some kind of information structural prominence.

¹³ $M \Rightarrow S(\text{Wh})$ is nearly equivalent to Grimshaw's (1997) constraint 'OpSpec'.

- b. $M \Rightarrow S(\text{Wh})$: If a wh-operator m_1 has scope over m_2 at M , then s_1 asymmetrically c-commands s_2 at S .
- c. $M \Rightarrow S(\text{Inf})$: If m_1 is [+prom] and m_2 is [-prom] at M , then s_1 asymmetrically c-commands s_2 at S .

The ranking in (15) prefers abstract syntax to reflect scope and to ignore information structure. Thus, the relative order of the arguments will be preserved, if they only differ in information structure, but it will not be preserved, if (only) a lower argument is a *wh*-element.

- (15) English ranking of $M \Rightarrow S$ constraints:
 $M \Rightarrow S(\text{Wh}) \gg M \Rightarrow S(\text{Arg}) \gg M \Rightarrow S(\text{Inf})$

It is easy to see that these three correspondence constraints potentially conflict: they compete for the *same means of encoding* for *different kinds of information*. As introduced above, I also assume constraints that govern the correspondence between M and P directly. These constraints are parallel to those in .

- (16) Constraints on $M \Rightarrow P$ mapping:
- a. $M \Rightarrow P(\text{Arg})$: If an argument m_1 is higher than another argument m_2 at M , then p_1 precedes p_2 at P .
 - b. $M \Rightarrow P(\text{Wh})$: If a wh-operator m_1 has scope over m_2 at M , then p_1 precedes p_2 at P .
 - c. $M \Rightarrow P(\text{Inf})$: If m_1 is [+prom] and m_2 is [-prom] at M , then p_1 precedes p_2 at P .

The ranking of these three constraints is not parallel to the ranking for $M \Rightarrow S$ mapping.¹⁴

- (17) English ranking of $M \Rightarrow P$ constraints:
 $M \Rightarrow P(\text{Inf}) \gg M \Rightarrow P(\text{Arg}) (M \Rightarrow P(\text{Wh}))$

Most crucially, $M \Rightarrow P(\text{Inf})$ is higher than $M \Rightarrow P(\text{Arg})$. Thus, although information structure cannot override argument structure in S , it can do so in P – prominent elements cannot be moved in S , but they can be fronted in the string. The cause for the two different fronting strategies that we find in English is that *wh*-elements are fronted in S , while discourse prominent phrases are fronted in P .

To front a non-subject in S , we need an additional projection that provides a specifier as the landing site. This is the motivation for ‘*do*’-support in non-subject *wh*-questions.¹⁵ These constraints together ensure that $M \Rightarrow P(\text{Wh})$ is fulfilled by winners. $M \Rightarrow P(\text{Wh})$ will be left out in the discussion below, but it certainly has a function in languages other than English. One could imagine that it is responsible for multiple *wh*-

¹⁴A brief note about the constraint $M \Rightarrow P(\text{Wh})$ in (17): Its actual ranking cannot be determined because its effects are hidden behind the high rank of $M \Rightarrow S(\text{Wh})$ and constraints on $S \Rightarrow P$ correspondence, in particular, $S \Rightarrow P(\text{sh})$, $S \Rightarrow P(\text{hc})$, and $S \Rightarrow P(\text{NP})$.

¹⁵I am abstracting away from the special problem of *do*-support here. For an OT-analysis, see Grimshaw (1997). This analysis should, however, be made compatible with what is proposed here. The insertion of “*do*” can be assumed to be forced by a highly ranked constraint that requires to keep the relative order of subject and main verb.

fronting in, for instance, many Slavic languages.¹⁶ The two different fronting strategies result from the ranking in (18) (the family of $S \Rightarrow P$ constraints is abbreviated with one general constraint here):

- (18) English ranking of $M \Rightarrow S$, $M \Rightarrow P$, and $S \Rightarrow P$ constraints:
 $M \Rightarrow S(Wh) \gg \{M \Rightarrow P(Inf), M \Rightarrow S(Arg)\} \gg S \Rightarrow P \gg \{M \Rightarrow S(Inf), M \Rightarrow P(Arg)\}$

Let us now take a look at the OT analyses. We will start with *wh-movement*. The constraint $M \Rightarrow S(Wh)$ is higher than $M \Rightarrow S(Arg)$, hence, a structure with syntactic *wh*-movement is preferred over one without. The effects of this movement are preserved at P. This is ensured by ranking $S \Rightarrow P$ higher than $M \Rightarrow P(Arg)$.

Let us start with a standard OT system with the following properties: we use M as input representation, a specification of all relevant semantic information, including at least argument structure, scope specification and information structure. Candidates are [S,P] pairs:

- (19) Standard OT syntax model:
 Input: M
 Output: S, P

We now want to derive (3b), repeated in (20):

- (20) *What did John write?*

For S, we have to consider two candidates, one with *wh*-movement (abbreviated as “+mvt”), and one without (“-mvt”):

- (21) Candidates for S:
 a. +mvt: [NP[+wh]_i do_j NP t_j [V t_i]]
 b. -mvt: [NP V NP[+wh]]

For P, we in principle have to consider all possible permutations of subject, (*do*), verb and object. I will restrict myself to the orders SVO, OSV and OVS, and parallel cases with *do* for the [+mvt] candidates. The other orders are ruled out independently. SOV is ruled out by the ranking “ $S \Rightarrow P(hc) \gg S \Rightarrow P(ch)$ ” which sets the head parameter to VO. Verb-initial orders are ruled out by $S \Rightarrow P(sh)$ which ensures that specifiers precede the heads of their projections.

- (22) Candidates for P:
 a. [-mvt]: SVO, OSV, OVS
 b. [+mvt]: *SdoVO*, *OSdoV*, *OdoSV*

This gives us six candidate [S,P] pairs. The following table only uses the constraints relevant for the discussion of *wh*-movement, tacitly assuming that information structure is not relevant in this particular case.

¹⁶See (Legendre, 2000) for an account of Bulgarian *wh*-clusters along similar lines.

(23) *Wh*-movement of an object in English:

Qx.write(j,x)	M⇒S (Wh)	M⇒S (Arg)	S⇒P (sh)	S⇒P (hc)	S⇒P (NP)	M⇒P (Arg)
[+mvt,SdoVO]		*	*!	*	*	
[+mvt,OSdoV]		*		*!		*
☞ [+mvt,OdoSV]		*				*
[-mvt,SVO]	*!					
[-mvt,OSV]	*!			*	*	*
[-mvt,OVS]	*!		*	*	*	*

The highest constraint is $M \Rightarrow S(Wh)$. So the three candidates without *wh*-movement at S are excluded. For the three candidates left, we have to find out the optimal linear order, they differ only at P. They all violate $M \Rightarrow S(Arg)$ the same way, because of the syntactic *wh*-movement. So this constraint cannot decide. The $S \Rightarrow P$ constraints¹⁷ determine as winner the *OdoSV* order: the elements are linearised according to their relative c-command relations (O asymmetrically c-commands *do*, which asymmetrically c-commands S, which asymmetrically c-commands V). Only the heads of movement chains are taken into account in the evaluation. $S \Rightarrow P(sh)$ needs to be fulfilled by O, the fronted *wh*-phrase, and *do*. $S \Rightarrow P(hc)$ can only be violated by *do* and its complement [S V]. $S \Rightarrow P(NP)$ is fulfilled, if O precedes S. For the [-mvt] candidates the violations are determined accordingly.

Let us now turn to (13c), left dislocation of a discourse prominent object, repeated in (24).

(24) *The red book, John wrote.*

We use the same candidates as before. The constraints on operator scope are left out, as they are not active here. The $S \Rightarrow P$ constraints have the general effect of keeping departures from the default linear order at P, relative to a given S, as *minimal* as possible. This is again decisive, as (25) shows:

(25) Topicalisation of an object in English:

write(j,rb) rb=[+PROM]	M⇒P(Inf)	M⇒S(Arg)	S⇒P (sh)	S⇒P (hc)	S⇒P(N P)	M⇒P(Arg)	M⇒S(I nf)
[+mvt,SdoVO]	*!	*	*	*	*		
[+mvt,OSdoV]		*!		*		*	
[+mvt,OdoSV]		*!				*	
[-mvt,SVO]	*!						*
☞ [-mvt,OSV]				*	*	*	*
[-mvt,OVS]			*!	*	*	*	*

The candidates with syntactic object movement are ruled out early because of the sub-ranking ‘ $M \Rightarrow S(Arg) \gg M \Rightarrow S(Inf)$ ’. $M \Rightarrow P(Inf)$ enforces linear reordering according to information structural needs, and the $S \Rightarrow P$ constraints choose the candidate that preserves most of the underlying syntactic configuration in the linear order. This is the candidate ‘[-mvt,OSV]’ where the object is left dislocated, which is, according to the analysis presented here, a process of mere linear reordering. The constraint $S \Rightarrow P(sh)$ makes the difference in the comparison with OVS order. The relative order of the head

¹⁷We use three of the constraints introduced informally in section 2. Their relative ranking cannot be determined with the data at hand. All rankings yield the same result.

V and its complement O cannot be preserved, the object is topicalised. But the OSV candidate preserves the relative order of V and its specifier, S.

4 Topicalisation, *Wh*-movement, and Verb-Second in German

This section discusses the equivalent German data. The clauses in (3) translate into German as follows:

- (26) a. *John schrieb dieses Buch*
John wrote this book
b. *Was schrieb John?*
What wrote John?
c. *Das rote Buch schrieb John*
The red book wrote John

At first sight, to account for these examples seems rather easy: If topicalisation has the same syntactic effects as *wh*-movement, then perhaps because $M \Rightarrow S(\text{Inf})$ is ranked as high as $M \Rightarrow S(\text{Wh})$:

- (27) Possible German ranking I:
 $\{M \Rightarrow S(\text{Wh}), M \Rightarrow S(\text{Inf})\} \gg M \Rightarrow S(\text{Arg}) \gg S \Rightarrow P \gg \{M \Rightarrow P(\text{Inf}), M \Rightarrow P(\text{Arg})\}$

Alternatively, one could yield the same effect in an indirect way: ranking both $M \Rightarrow P(\text{Inf})$ and $S \Rightarrow P$ higher than $M \Rightarrow S(\text{Arg})$ has the effect that a structure with object movement at S is preferred in order to make S and P maximally correspond:

- (28) Possible German ranking II:
 $\{M \Rightarrow S(\text{Wh}), M \Rightarrow P(\text{Inf})\} \gg S \Rightarrow P \gg M \Rightarrow S(\text{Arg}) \gg M \Rightarrow P(\text{Arg})$

However, German is a head-final language, with the exception of the Verb-Second effect. To account for this, we need to be more detailed about the constraint family $S \Rightarrow P$. We already introduced a number of these constraints in section 2 informally. (29)-(31) give more precise definitions:

- (29) a. **$S \Rightarrow P(\text{ch})$** – ‘complement before head’:
If s_1 and s_2 are sister nodes at S, and s_1 is a head and s_2 its complement, then p_1 precedes p_2 at P.
b. **$S \Rightarrow P(\text{hc})$** – ‘head before complement’:
If s_1 and s_2 are sister nodes at S, and s_1 is a head and s_2 its complement, then p_1 precedes p_2 at P.

The relative ranking of $S \Rightarrow P(\text{hc})$ and $S \Rightarrow P(\text{ch})$ determines whether a language is head-initial ($S \Rightarrow P(\text{hc}) \gg S \Rightarrow P(\text{ch})$) or head-final ($S \Rightarrow P(\text{ch}) \gg S \Rightarrow P(\text{hc})$). The two constraints in (30) express default linearisation tendencies for further relations within a phrase.

- (30) a. **$S \Rightarrow P(\text{sa})$** – ‘sisters are adjacent’:
If s_1 and s_2 are sisters at S, then p_1 and p_2 are adjacent at P.

- b. **S \Rightarrow P(sh)** – ‘specifier precedes head’:
 If s_1 is the specifier of a maximal projection headed by s_2 at S, then p_1
 precedes p_2 at P.

The two constraints in (31) restrict shape conservation to syntactic categories: Asymmetric c-command between elements of the same syntactic category translates into precedence.

- (31) a. **S \Rightarrow P(NP)**:
 If an NP s_1 asymmetrically c-commands another NP s_2 at S, then p_1
 precedes p_2 at P.
 b. **S \Rightarrow P(V⁰)**:
 If a verbal head s_1 asymmetrically c-commands another verbal head s_2 at S,
 then p_1 precedes p_2 at P.

S \Rightarrow P(NP) already played an important role in the discussion of Greenberg’s first universal in section 2. S \Rightarrow P(V⁰) has been used by Schmid and Vogel (2004) in accounting for the dialectal typology of three-verb clusters in German.

Swiss German dialects, like Dutch dialects, display a default order in verbal complexes where the higher verb precedes the lower verb, while standard German displays the opposite order, obeying S \Rightarrow P(ch) even within verbal complexes. While Swiss German dialects are still head-final for the relative order of object and verb, complex V-VP structures are head-initial. The high-ranking of S \Rightarrow P(V⁰) introduces this kind of restricted head initiality into a verb-final language.

The two constraints have no counterpart constraint requiring a right-to-left order. This asymmetry reflects what is also expressed in Kayne’s (1994) *Linear Correspondence Axiom*, namely, a universal tendency for the iconic mapping of asymmetric c-command into precedence. Symmetric c-command, as observed in the case of head-complement sisterhood, is string-ambiguous, and this motivates the two constraints on head-complement order which directly implement a syntactic parameter. Specifiers again asymmetrically c-command their heads, so there is only a constraint that requires specifiers to precede heads.

The restriction to a particular syntactic category in S \Rightarrow P(V⁰) and S \Rightarrow P(NP) might sound arbitrary. It would do less so, if we could turn it into a restriction on the formulation of constraints: constraints on the relative linear order of syntactic elements relate either the immediate constituents of a phrase (specifier, head, complement) or elements of the same syntactic category. I will follow this restriction in this paper. It is a mode of implementing the idea of *relativised* minimality (cf. Rizzi 1990): only like elements can block each other by minimality.

One could imagine that relative prominence is also important for the head-complement relation: though the two elements are sisters, the head could be seen as relatively more prominent, because, in traditional terminology, it often governs and selects its complement. Thus, although the structural symmetry of the head-complement relation justifies the two mirror image constraints S \Rightarrow P(ch) and S \Rightarrow P(hc), under extreme conditions head-before-complement might emerge as the unmarked case.¹⁸

The verb-second effect in German could be interpreted as the consequence of such a situation. German is head-final, except for the highest projection in the clause. For the highest projection, the head parameter has the biggest consequences. Its head is either

¹⁸In Schmid and Vogel (2004), we discuss another example in case, namely the loss of strict complement-head order in standard German verbal complexes with rising complexity.

the first (given that the specifier is empty) or the last element of the clause. Let us assume a constraint that reflects this:

- (32) **S \Rightarrow P(hc-top):**
 If s_1 is the head of the topmost projection in S, and s_2 its complement, then p_1 precedes p_2 at P.

High rank of this constraint turns an OV language like German into a ‘context sensitive VO-language’. The highest head of an extended projection is more important than other heads because it signals the categorial status of the phrase. If this is the motivation for S \Rightarrow P(hc-top), then there should not be a counterpart constraint S \Rightarrow P(ch-top) requiring the topmost head to occur at the end of the clause. In an SOV language, the effects of such a constraint could not be seen, but for an SVO language this constraint would predict a verb-last effect for the highest verb. Thus, while the German order of a matrix clause with an analytic tense is ‘S Aux O V’, the imaginary language we are talking about would have ‘S V O Aux’, and a clause with simple tense would have ‘S O V’. As far as I know, such a language does not exist. I take this gap as another piece of evidence for the claim made above that precedence is the *iconic* translation of relative syntactic prominence into relative linear order. The standard German ranking, then, is:

- (33) S \Rightarrow P(hc-top) >> S \Rightarrow P(ch) >> S \Rightarrow P(hc)

Let us now return to our discussion of topicalisation and *wh*-movement in German. The main difference to English is that German does not have the English type topicalisation. Fronting is always accompanied by verb-second. One ranking option is that German has M \Rightarrow S(Inf) ranked higher than M \Rightarrow S(Arg). Linear reordering at P would then be unnecessary to fulfil M \Rightarrow P(Inf). However, left dislocation is also impossible in the case of fronted adverbials, cf. the following three clauses:

- (34) a. *Yesterday I met Pierce Brosnan*
 b. **Gestern ich traf Pierce Brosnan*
 yesterday I met P. B.
 c. *Gestern traf ich Pierce Brosnan*

Adverbials do not need to undergo movement, they can be inserted where they occur at the surface. There is nothing so far in the system of constraints that would prevent (34b). The ban on left dislocation of the English kind is more general. German has a construction that looks very much like left dislocation, but here it is crucial that the left dislocated element has a correlate within the matrix clause:

- (35) a. *Gestern, da traf ich Pierce Brosnan*
 yesterday there met I P. B.
 b. *(Der/Den) Pierce Brosnan, den traf ich gestern*
 (the-NOM/the-ACC) P. B. that-one-ACC met I yesterday

I assume that the clause that follows the left dislocated element must contain all its constituents, and, furthermore, that its *shape* has to be as would be expected from S \Rightarrow P mapping constraints, and, in particular, the *left edge of P* has to correspond to the *highest constituent in S*. Remember that the analysis of English left dislocation

presented in section 3 assumes a *mismatch* between S and P: the leftmost element in P is the fronted constituent. It is only in initial position in P, not in S, so the highest constituent in S is another element, usually the subject. I assume that German has a ban on such a mismatch in the clause-initial position. We already saw in our discussion of the shift from head-final to head-initial in the highest clausal projection that the initial projection of a clause has a special status in German. The constraint that I propose is the one in (36):

- (36) **S \Rightarrow P(top):**
 If s_1 is the highest constituent in S, then p_1 is leftmost in P

The notion of ‘highest constituent’ is defined as follows:

- (37) **HIGHEST CONSTITUENT:**
 Given a syntactic structure S with the root node R, α is the highest constituent of R, iff α is an X^0 or XP node dominated by R, and there is no other X^0 or XP node $\beta \neq R$, such that β dominates or asymmetrically c-commands α .

A complementiser-introduced subordinate clause now has two highest constituents, because the complementiser and its complement mutually c-command each other. S \Rightarrow P(top) can only be fulfilled for one of them. But S \Rightarrow P(hc-top) ensures that the complementiser is the first element. The fronting of a third constituent would be dispreferred, as it results in a double violation of S \Rightarrow P(top).

For the examples in (35), we can assume that the left dislocated phrase is outside of the clause in both S and P. The highest constituents of the matrix clauses are the resumptive pronouns ‘*da*’, ‘*der/den*’, respectively. S \Rightarrow P(top) is fulfilled. With high rank of this constraint, we only need to have M \Rightarrow S(Inf) or M \Rightarrow P(Inf) ranked higher than both M \Rightarrow S(Arg) and M \Rightarrow P(Arg).¹⁹

Let us now discuss some concrete examples. First, a simple subject-initial clause:

- (38) *Hans liebt Maria*
 H.-NOM loves M.-ACC

We keep the syntactic structure constant for the candidates, a simple VP which is represented according to X-bar theory, and we only consider the six logically possible orders of the elements. The constraints in the tableau are those that are relevant here:

¹⁹The decision between these two options is relevant for the treatment of scrambling. If scrambling is adjunction, then we have M \Rightarrow S(Inf) ranked lower than M \Rightarrow S(Arg) and M \Rightarrow P(Arg) (and M \Rightarrow P(Inf) ranked higher), if scrambling is movement to a specifier position, then the constraint that triggers it, for instance M \Rightarrow S(Inf), is ranked higher. S \Rightarrow P(top) ensures that S and P are synchronised in their initial constituent. The standard analysis of scrambling is that it is adjunction, and so we may assume that M \Rightarrow S(Inf) is ranked low. I will assume this in the subsequent discussion.

(39) Simple clause structure in German:

[VP S [v' V O]]	S⇒P (top)	S⇒P (hc-top)	S⇒P (ch)
☞ SVO			*
SOV		*!	
VOS	*! *		*
OVS	*! *	*	
VSO	*!		*
OSV	*!	*	

With an auxiliary, the picture slightly changes: the auxiliary is in second position, and the main verb is clause-final:

(40) *Hans hat Maria geküsst*
H.-NOM has M.-ACC kissed

S⇒P(hc-top) is fulfilled by the auxiliary, and so the verbal participle occurs clause-finally in order to fulfil S⇒P(ch). For reasons of space, we only display candidates with SOV and SVO order in the tableau below:

(41) Simple clause structure with auxiliary in German:

[IP S Aux [VP [v' V O]]]	S⇒P (top)	S⇒P (hc-top)	S⇒P (ch)
c1: AuxSVO	*!		**
c2: AuxSOV	*!		*
c3: SAuxVO			**!
☞ c4: SAuxOV			*
c5: SVAuxO		*!	*
c6: SOAuxV		*!	*
c7: SVOAux		*!	*
c8: SOVAux		*!	

We see that the two constraints that take care of the topmost projection exclude all structures that do not begin with the sequence “S Aux”. S⇒P(ch) then makes the decision in favour of the candidate c4 which has head final order for the VP.

I leave out the table that shows the result for subordinate clauses. The winning candidate roughly conforms to c8 in the previous tableau, with the addition of an initial complementiser. That structure performs optimally on S⇒P(top) and S⇒P(hc-top) because of the complementiser, and has the ideal head-final order for the non-initial projections, which leads to a minimal violation of S⇒P(ch). Next, we turn to an example with a topicalised direct object:

(42) *Das Buch schrieb Helga*
the book wrote H.

(43) Topicalisation of an object in German:

	S \Rightarrow P (top)	S \Rightarrow P (hc-top)	M \Rightarrow P (Inf)	S \Rightarrow P (ch)
☞ schrieb(h,b) b=[+PROM]				
☞ [+mvt,OVS]				*
[-mvt,SVO]			*!	
[+mvt,OSV]		*!		*
[-mvt,SOV]		*!	*	

A number of further possible candidates are not listed here, for instance, V-initial orders. As all of these alternative variants would violate S \Rightarrow P(top), I excluded them here for ease of representation. We only consider candidates that fulfil this constraint. The second constraint, S \Rightarrow P(hc-top), enforces verb-second and excludes the candidates with the verb in final position. Between the two remaining candidates, SVO order without movement, and OVS order with movement, M \Rightarrow P(Inf) makes the decision. This reconstructs the standard analysis of such cases: object fronting is triggered by information structure, thereby observing the verb-second requirement in German.

The case of wh-movement is very similar. Here, the constraint M \Rightarrow S(Wh) makes the decision:

(44) *Was schrieb Helga?*
What wrote H.

(45) Wh-movement of an object in German:

Qx.schrieb(h,x)	S \Rightarrow P (top)	S \Rightarrow P (hc-top)	M \Rightarrow S (Wh)	S \Rightarrow P (ch)
☞ [+mvt,OVS]	*			*
[-mvt,SVO]			*!	
[+mvt,OSV]		*!		
[-mvt,SOV]		*!	*	

5 Multiple Questions

The system developed thus far predicts minimality effects with multiple questions. Consider the following German example:

(46) *Was kaufte wer?*
what bought who?

With two *wh*-phrases, the problem arises that only one of them can fulfil M \Rightarrow S(Wh). We would either have a structure like (47a) or (47b), in both of which the verb is outside of the scope of the lower *wh*-phrase.

- (47) a. [VP [NP *wer*] [V' [V⁰ *schrieb*] [NP *was*]]]
b. [VP [NP *was*]_i [V' [V⁰ *schrieb*]_j [VP [NP *wer*] [V' t_j t_i]]]]

As a consequence of this, M \Rightarrow S(Wh) cannot decide between the two structures. M \Rightarrow P(Wh) could be fulfilled by a candidate that has both *wh*-phrases in front in the linear string, but this would violate the highly ranked S \Rightarrow P(hc-top) which requires the

complement of I^0 , including the lower *wh*-phrase, to remain to the right of the finite verb. In addition, structure (47b) performs worse than (47a) in $M \Rightarrow S(\text{Arg})$ and $M \Rightarrow P(\text{Arg})$. Therefore, whatever constraint ranking one chooses, (47b) will lose against (47a). In the light of an example like (46), this might look like a bad result. On the other hand, this situation simply derives the superiority effect. For English where a clause like (46) is ungrammatical this would be a fine result, and even German native speakers often judge clauses like (47b) worse than (47a). There is a universal markedness relation between these two structures. Probably all languages that allow for (47b) also allow for (47a), but not vice versa.

The survival of (46) in German in spite of its markedness must be due to another factor. The standard OT answer to such situations is that there is not only markedness, but also faithfulness – the force to preserve the input in the output. In the system developed thus far, there is no place for such faithfulness. The correspondence constraints that we established are in fact a kind of faithfulness constraints under the more general conception of faithfulness and correspondence developed by McCarthy and Prince (1995). But our correspondence constraints only evaluate the translations between different types of representations. (47a,b) are of equal status here, and it is impossible to give (47b) the advantage it needs. This would only be possible, if we specified in the input already that we want structure (47b).

Thus far, I assumed a standard version of OT syntax with M in the input, and pairs $[S,P]$ as output candidates. What we obviously need, is a specification for S in the input already, and faithfulness constraints controlling for the preservation of the ‘input S ’ in the ‘output S ’. I argued for such an architecture elsewhere already.²⁰

The main motivation for this move is the observation that there usually is more than one way to express the same meaning. Different modes of expression cannot always be traced back to some functional or subtle semantic-pragmatic difference in a predictable manner. The choice of syntactic construction in an utterance is often guided by stylistic and rhetorical factors, perhaps statistical factors like salience, and could even be simply accidental. The alternative to specifying S in the input would be an integration of any such factors into M , with the underlying assumption that all syntactic differences *must* have a predictable semantic or pragmatic ‘cause’. We are far from being able to decide whether this is true or not. I assume that it is *not* true. Anyway, for now it seems more practicable to *presuppose* this decision for a particular syntactic structure. The OT grammar then evaluates whether that structure – represented by the syntactic part of the input – is well-formed, i.e., is correlated with a faithful output candidate by the evaluation procedure.

One reviewer expresses a worry concerning this architecture. According to him, it ‘raises the clear danger of stipulating in the input all the syntactic properties that we wish to derive and then use faithfulness to distinguish one structure from the other. The result would be highly stipulative with no genuine explanation of the syntactic properties at issue.’ To my mind, this worry does not affect the current approach in any sense.

First of all, the question of what is a legitimate syntactic structure in general is not decided casewise for individual languages in generative grammar. A legitimate syntactic structure is any structure that can be built with the candidate generation function, GEN. In this paper, I assume for GEN a simplified version of X-bar theory, excluding adjunction, and allowing for ‘forests’ of phrases instead. All we need is a well-defined grammar formalism. Whether that formalism is stipulative in itself is an

²⁰See Vogel (2002, 2004). Baković and Keer (2001) were the first to propose such an input for OT syntax.

independent issue.

The question which of these structures is possible in a language is answered by that language's grammar, its constraint hierarchy. The constraint system that I use here is also very restricted and independently motivated by the underlying concept of correspondence. It is not the task of the constraint hierarchy to explain the syntactic properties of winning candidates. This is again the subject of GEN. The constraint hierarchy explains why candidates with such properties are grammatical in a given language. The reviewer's worry concerns a possible unsystematic usage of faithfulness. I think that this danger is much higher in an alternative to the proposed system that has only M in the input.

I showed above that the proposed system predicts a typological implication, namely, that languages that allow for object-subject orders also allow for subject-object orders, but not necessarily vice versa. The ability to derive such markedness relations is a well-known property of OT systems using faithfulness.

In order to account for optionality, the alternative theory has to play around with the definition of M in a very stipulative manner. The risk of losing typological implications because of arbitrary input definitions appears to me much higher.

The whole discussion boils down to one central question which is answered in two different ways, the problem of syntactic optionality. The two answers are:

1. Optionality of syntactic structures is an irreducible property of natural languages. Therefore the input must be a pair [M,S].
2. Optionality of syntactic structures is only apparent. There are no two syntactic structures in any language of the world that have the same meaning. Therefore the input must contain M only.

To my mind, answer 2 is nothing more but wishful thinking, although it might be correct for a subset of cases of syntactic optionality. A functional argument from language use in favour of answer 1 is that answer 2 requires a level of linguistic precision that is never reached in everyday conversation. Language is an efficient means for communication because of this robustness. We can understand each other although the expressions we use are often 'imperfect'.

The alternative strategy, suggested by the reviewer, asks the question 'How is a meaning M expressed in a language? ', on which the model proposed here comments that there are many ways of expressing M, and that it is impossible to give a single answer. Therefore the question has to be restated: 'Is it possible to express a meaning M with the structure S in a language? '. This strategy presupposes a certain amount of imperfection and non-correspondence. It describes the degree of tolerance of such imperfections by the relative ranking of faithfulness. Note also that this is the strategy that is typically chosen when linguists do syntactic typology. We ask questions like 'Is it possible to have superiority violations in multiple questions in a language?'. This question describes an input that contains both a semantic (multiple question) and a syntactic (the object *wh*-phrase precedes the subject *wh*-phrase) specification.

To sum up, the model we arrive at is sketched in (48), input and output S are distinguished by subscripts.

- (48) Input and output representations in OT syntax (Vogel 2002, 2004):
Input: S_I, M
Output: S_O, P

Preservation of S_I in S_O is enforced by faithfulness constraints. We could assume a large set of such faithfulness constraints, but for our purpose it is sufficient to use the general formulation in (49):

- (49) $S_I \Rightarrow S_O$:
 S_I and S_O are identical.

To ensure that the minimality violating structure survives, we only need to rank $S_I \Rightarrow S_O$ higher than $M \Rightarrow S(\text{Arg})$, and we get the correct predictions for multiple questions in German.²¹

6 Freezing, Recoverability and Bidirectionality

Free word order languages sometimes display surprising minimality effects with ordinary NPs, under surface ambiguity of the NPs involved. This phenomenon of “word order freezing” has been discussed at length by Lee (2001*a*, 2001*b*, this volume). In colloquial Korean, case markers can often be omitted. However, the NPs are now surface ambiguous for case, and their order is no longer free, subject-object order is the only available option:

- (50) *Mary Jane manna-ss-e*
 Mary Jane meet-pst-decl
 a. ‘Mary met Jane.’
 b. *‘Jane met Mary.’
 (Lee 2001*b*, 111)

It is not surprising that German, another free word order language, shows the same effect:

- (51) German, freezing to SO order :
 a. *Den HANS liebt Maria, ohne ihn zu kennen*
 the H.-ACC loves M. without him-ACC to know
 ‘It is Hans who Maria loves, without knowing him’
 b. *HANS liebt Maria, ohne ihn zu kennen*
 H.-NOM/ACC loves M.-ACC/NOM without him-ACC to know
 ?? ‘It is Hans who loves Maria without knowing him’

(51b) has a strange interpretation, the person called ‘*Maria*’ is male. The reason is obviously that the plausible interpretation with the underlying marked OVS order in the matrix clause is inaccessible. Both proper nouns are morphologically unspecified for case.

The important observation here is that a syntactic constraint, in particular: $M \Rightarrow S(\text{Arg})$ is active in the interpretive optimisation. We have a minimality effect in interpretation. This suggests that the two perspectives of optimisation are two sides of the same coin. We can combine them in a bidirectional model of OT syntax that uses

²¹A well-known fact about English multiple questions is that discourse-linked *wh*-phrases can remain in situ. A clause like “*What did which student read?*” is grammatical. I sketch an analysis of these cases within a correspondence-based OT syntax in Vogel (2004).

the same set of ranked constraints for production-oriented and interpretation-oriented optimisation. I follow Lee in this conclusion.

The two perspectives are combined in the definition of grammaticality given below in (53). It implements Pesetsky's (1997, 1998) intuition that recoverability of underlying information constrains surface forms. Contrary to Pesetsky's original approach, recoverability is not taken to be an OT constraint, but a condition on grammaticality and therefore inviolable. In a model that generates [underlying form, surface form] pairs, a violation of recoverability by a given pair proves its illegitimacy. If u_1 is not recoverable from s_1 , then a different underlying form must be more optimal for s_1 , e.g., u_2 . In such a case, u_1 is blocked by u_2 in the interpretive optimisation for the input s_1 .

I use the terms "first" and "feedback" optimisation for the two optimisation steps. This suggests a serial order of the two processes, but this is not essential. A reversal in the order would not change the results. The two optimisation steps should be seen as operating in parallel. Lee (2001a, 2001b, this volume) uses the terms "productive and interpretive optimisation", following Smolensky (1996). The reason why I do not adopt these terms is that I want to avoid the usage of vocabulary from performance grammar when discussing the properties of a competence grammar. The two optimisation processes can be seen as independent of each other. They are only combined because of the way we define grammaticality.

(52) Input and output representations in bidirectional OT syntax:

First optimisation:	Input: S_i, M
	Output: S_o, P
Feedback optimisation:	Input: P
	Output: S_i, M

(53) **Grammaticality:**

A triple $[M_i, S_i, P_i]$ is grammatical, if and only if the input $[M_i, S_i]$ yields $[S_i, P_i]$ in first optimisation, and the input $[P_i]$ yields $[M_i, S_i]$ in feedback optimisation.

Instead of repeating Lee's discussion here, I want to introduce additional observations concerning the behaviour of German *wh*-phrases. While a clause with two ambiguous *wh*-phrases like (54a) shows the same freezing effects as the examples given in (51), the effect disappears if only one of the two ambiguous NPs is a *wh*-phrase (54b,c):

- (54) a. *Welche Lehrerin besuchte welches Kind, ohne sie zu kennen?*
 which teacher-FEM-NOM/ACC visited which child-NEUT-NOM/ACC without her-ACC to know
 inaccessible interpretation:
 "Which teacher did which child visit without knowing her?"
- b. *Welche Frau küsste Hans, ohne sie zu kennen?*
 which woman-NOM/ACC kissed H. without her to know
 "Which woman did Hans kiss without knowing her?"
- c. *Welche Frau küsste Hans, ohne ihn zu kennen?*
 which woman-NOM/ACC kissed H. without him to know
 "Which woman kissed Hans without knowing him?"

This lacking of a freezing effect is unexpected under the account as presented thus

far. In the input, we have a marked structure, an OVS order. High ranked syntactic faithfulness, $S_I \Rightarrow S_O$, ensures that this order survives first optimisation although it violates $M \Rightarrow S(\text{Arg})$. In feedback optimisation we have the P part of the winning candidate as input. Freezing occurs if P is morphologically ambiguous for two underlying S's, the marked OVS and the less marked SVO. This is the case here. $S_I \Rightarrow S_O$ can no longer rescue the marked structure, there is no S representation in the input, and so $M \Rightarrow S(\text{Arg})$ becomes decisive in feedback optimisation, and SVO wins against OVS, the criterion for grammaticality given above is not met. The tables in (55) illustrate this.

(55)

First optimisation:		
	$S_I \Rightarrow S_O$	$M \Rightarrow S(\text{Arg})$
S= OVS		
☞ OVS		*
SVO	*!	

Feedback optimisation:		
	$S_I \Rightarrow S_O$	$M \Rightarrow S(\text{Arg})$
P= <i>Hans...Maria</i>		
OVS		*!
☞ SVO		

With disambiguation by the preceding determiner ‘den’ in (51a), the marked structure survives because $S \Rightarrow P(\text{case})$ is ranked higher than $M \Rightarrow S(\text{Arg})$ and enforces the noun with accusative morphology to be interpreted as direct object. The scenario in (55) holds for (54a) and (54c), but it makes the wrong predictions for (54b). Obviously, argument structure makes no decision at all here, OVS is not blocked by SVO. It seems as if there was no constraint $M \Rightarrow S(\text{Arg})$ at all. But we know that it is there.

What makes the difference is, of course, the fact that we are comparing the relative order of a *wh*-NP and a non-*wh*-NP. The two do not interfere with each other. This observation is in line with Rizzi’s (1990) conception of *relativised minimality*.

But such relativisation is not encoded in our constraint set yet. An easy way of doing it would be assuming that $M \Rightarrow S(\text{Arg})$ applies only to XPs of the same type. In that case, $M \Rightarrow S(\text{Arg})$ could not be violated, because the two NPs are of different type. This would work quite well with German, but we would run into the danger of being unable to derive *wh*-in situ languages like Chinese.

The sub-ranking ‘ $M \Rightarrow S(\text{Arg}) \gg M \Rightarrow S(\text{Wh})$ ’ is a reasonable candidate for the driving force behind *wh*-in situ. Hence, $M \Rightarrow S(\text{Arg})$ must be violated by *wh*-fronting in a clause like (54b). But it may not be violated by the same structure in German, as the grammaticality of this structure shows.

As a way out of this dilemma I propose that the argument hierarchy is derived differently in the two languages. This difference results in different behaviour of the same structure with respect to $M \Rightarrow S(\text{Arg})$. Consider a predicate logic formula for a simple question like “*who did John kiss?*” as part of its semantic representation, M:

(56) $Qy.kiss(x,y)$

For the determination of the argument hierarchy on the basis of this formula, it is crucial which occurrence of ‘y’ in (56) is decisive. Let us assume that this decision is made differently in different languages. In German the argument hierarchy is based first

on the operatorhood of an element and in the second place on the argument position, while in Chinese it is the other way around. This can be described in an OT fashion:

(57) Determination of argument hierarchies in German and Chinese:

German:		
Qy.kiss(x,y)	Operator	Arg. Str.
x > y	*!	
☞ y > x		*

Chinese:		
Qy.kiss(x,y)	Arg. Str.	Operator
☞ x > y		*
y > x	*!	

The argument hierarchies are presupposed by the constraints $M \Rightarrow S(\text{Arg})$ and $M \Rightarrow P(\text{Arg})$. *Wh*-in-situ now fulfils $M \Rightarrow S(\text{Arg})$ in Chinese, while it violates that constraint in German for the given structure. However, when the *wh*-phrase is fronted, $M \Rightarrow S(\text{Arg})$ is fulfilled in German, no matter whether the *wh*-phrase is subject or object, provided that it is the only *wh*-phrase in the clause. In Chinese, $M \Rightarrow S(\text{Arg})$ can only be fulfilled if the *wh*-phrase remains in situ. Ranking $M \Rightarrow S(\text{Arg})$ higher than $M \Rightarrow S(\text{Wh})$ derives the Chinese pattern.

I finally want to show that the bidirectional strategy can successfully be applied to a long-standing empirical problem in German syntax which is exemplified in (58).

- (58) a. Wem₁ hat er abgeraten, sofort wem₂ nach Saloniki nachzureisen?
 who-DAT has he dissuaded immediately who-DAT to Saloniki to-travel-after
 ‘Who has he dissuaded from travelling after whom to Saloniki?’ ”
- b. *Wem_{1/2} hat er (t₁) wem_{2/1} abgeraten, sofort t₂ nach Saloniki nachzureisen?
 who-DAT has he who-DAT dissuaded immediately to Saloniki to-travel-after
 intended reading: ‘Who is the person such that he has dissuaded whom to travel after that person?’ ”
- c. Wohin₂ hat er wem₁ abgeraten, der Prinzessin t₂ nachzureisen?
 where-to has he who-DAT dissuaded the princess-DAT to-travel-after
 ‘What is the place such that he has dissuaded whom to travel there after the princess?’ ”
 (cf. Haider 1996, Haider 2000, Fanselow 1991)

The ungrammaticality of (58b) can be interpreted as a freezing effect across clause boundaries. We see in (58c) that an embedded *wh*-phrase may extract across a *wh*-phrase in the matrix clause. However, under surface identity of the two phrases, this option breaks down. The explanation for this problem follows the same line as before: In feedback optimisation, an underlying syntactic structure is chosen that fulfils $M \Rightarrow S(\text{Arg})$.²²

²²This example shows again, like the discussion we had before, that the determination of an argument hierarchy depends on the whole representation M, and cannot be restricted to argument

The difference to the freezing cases discussed above is that this winning interpretation is itself ungrammatical. This is due to the vacuous movement of the embedded *wh*-phrase into the middle field of the matrix clause. For the representation of this vacuous movement, we have two options in our model. It could be represented at both S and P, or, as adjunction, only at P. The latter option is the more realistic one, as the example looks like a typical case of scrambling, which is mostly, though not always, analysed as an instance of adjunction in German.²³

Let us first discuss how a candidate with vacuous syntactic movement of ‘*wem₂*’ in S is excluded. The structure of this candidate is sketched in (59):

(59) [CP *wh*-NP₁ ... *wh*-NP₂ t₁ V [CP ... t₂ V]]

To exclude this structure, we need a constraint that is ranked higher than $S_I \Rightarrow S_O$. Let us assume a constraint on ‘clause-mateness’ in the following way:²⁴

(60) **M \Rightarrow S(cm)**:

If a predicate and one of its arguments are both constituents of the same tree in S, then they must be clause-mates.

Note the conditional clause in this definition. In the case of adjunct arguments, verb and argument are not clause-mates in S. But we do not want the constraint to penalise such cases, because this is not a case of syntactic dislocation. $M \Rightarrow S(cm)$ can be violated by a winner in German, as exemplified by (58c). So the constraint must be ranked relatively low. It may, however, only be violated in order to fulfil a higher ranked constraint, like, e.g., $M \Rightarrow S(Wh)$. Thus, the constraint on vacuous movement we are looking for should be a conjoined constraint, penalising a violation of $M \Rightarrow S(cm)$ without fulfilling $M \Rightarrow S(Wh)$:

(61) **M \Rightarrow S(cm) &_{XP} M \Rightarrow S(Wh)**(*_{VAC}):

No simultaneous violation of $M \Rightarrow S(cm)$ and $M \Rightarrow S(Wh)$ by the same XP.

The ranking that we then need is the following:

(62) *_{VAC} >> $S_I \Rightarrow S_O$ >> $M \Rightarrow S(cm)$

This ranking excludes the structure in (59) already in first optimisation. The second option for deriving (58b) is interpreting it only as reordering at P. The S of this candidate would have the structure of (58a). Such an order would violate the following constraint:

hierarchies provided by the argument structures of individual lexical items.

²³In Vogel and Steinbach (1998), we show that dative objects in German share many syntactic properties with adjoined categories and contrast with NPs in A-positions, like accusative objects and subjects. Also from this perspective, an adjunction analysis is more realistic.

²⁴The definition of clause-mateness should rely on the notion of ‘extended projection’ in the sense of Grimshaw (1991), which roughly conforms to the array of VP, IP and CP nodes on top of the same lexical V. Two elements are clause-mates if they are dominated by the same extended projections of V. At P, we need to refer to the P correspondent of that extended projection, i.e. a particular substring of P. This has to contain the P correspondents of the two elements in question.

- (63) **M⇒P(cm):**
A predicate and its arguments are clause-mates at P.

One could also imagine a clause-mate constraint for S⇒P mapping:

- (64) **S⇒P(cm):**
If s_1 is a constituent of the clause s_2 at S, then p_1 is contained in p_2 at P.

A high rank of this constraint would result in the blocking of all P movement across clause boundaries. This might be a plausible restriction for German. However, the other constraints on S⇒P mapping already cover effects of such a constraint.²⁵ For this reason, I will avoid using it here.

The easy task is showing that (58a) has the ideal linearisation of elements for an underlying syntactic configuration where the matrix clause ‘*wem*’ is fronted:

- (65) First optimisation for (58a):

S=[<i>wem</i> ₁ ... <i>t</i> ₁ V [<i>wem</i> ₂ V]]	M⇒S (Wh)	M⇒S (Arg)	M⇒P (cm)
☞ <i>c1</i> : <i>wem</i> ... <i>wem</i> ... V	*		
<i>c2</i> : <i>wem</i> ... <i>wem</i> ... V ... V	*		*

Candidate *c2*, representing (58b), violates M⇒P(cm) in addition to M⇒S(Wh) (which is violated by both candidates) without gaining anything. *c1* therefore blocks *c2* in first optimisation. (58b) cannot have the interpretation given in the input in (65). However, under the interpretation where the embedded *wem* is fronted, first optimisation lets this order survive in the same way as it lets (58c) survive. In feedback optimisation, we have this order in the input, and look for its optimal underlying structure. We have three candidates, and we need to derive that the minimality violating structure, the original input (indicated with “⇒” in (66)) now loses:

- (66) Feedback optimisation of (58b):

P= <i>wem</i> ... <i>wem</i> ... V ... V	* _{VAC}	M⇒S(Wh)	M⇒S(Arg)	M⇒S(cm)	M⇒P(cm)
⇒ <i>c1</i> : [<i>wem</i> ₂ ... <i>wem</i> ₁ V [<i>t</i> ₂ V]]		*	*!	*	*
☞ <i>c2</i> : [<i>wem</i> ₁ ... <i>t</i> ₁ V [<i>wem</i> ₂ V]]		*			*
<i>c3</i> : [<i>wem</i> ₁ ... <i>wem</i> ₂ <i>t</i> ₁ V [<i>t</i> ₂ V]]	*!	*		*	*

The worst candidate is *c3*, the one with vacuous movement in S. Because of the high rank of *_{VAC}, this candidate has no chance in first optimisation either, as discussed above.²⁶ M⇒S(Wh) cannot be fulfilled by the candidates, because we have two *wh*-phrases and only one position where they could fulfil that constraint. And now it is again M⇒S(Arg), the minimality constraint, that makes the decision against the initial input. The winning underlying structure is one that has no chance to win with the P under consideration here, because of the availability of the order (58a) in first optimisation.

²⁵For example, S⇒P(hc) requires all elements of the complement to follow the head. Material that is moved in front of the head in P results in a violation of that constraint. Likewise, the condition on sister adjacency is violated in such a configuration.

²⁶Note that *c1* does not violate *_{VAC}, because its violations of M⇒S(cm) and M⇒S(Wh) come from two different XPs!

Clause (58b) is now excluded under all three possible inputs. The MLC obeying interpretation is excluded as an instance of vacuous movement in first optimisation. The MLC violating interpretation survives first optimisation, but loses in feedback optimisation against the MLC obeying structure, but this structure is paired in first optimisation with a more optimal linearisation, the one without vacuous movement in (58a). There is no grammatical triple [M,S,P] with (58b) as P according to the definition of grammaticality in (53).

How can the blocking interactions described here be avoided in the case of the well-formed (58c)? The difference to (58b) is that a directional phrase is extracted from the subordinate infinitival clause. The blocking candidate would be an interpretation where the directional PP belongs to the matrix verb, and the vacuously moved dative object would belong to the embedded verb. This would be equivalent to a clause like (67):

- (67). **Wohin*₂ hat er abgeraten, *wem*₁ der Prinzessin *t*₁ nachzureisen?
 where-to has he dissuaded who-DAT the princess-DAT to-travel-after
 “Whereto did he dissuade to travel after whom after the princess?”

While one argument of ‘*dissuade*’, the addressee, is missing, the verb has an additional argument, the directional PP, of which we cannot make sense. Likewise, the embedded verb now has two dative objects instead of one, the thematic role of one of them is unclear. This candidate is subject to semantic markedness in a way that (58b) is not, because in the latter case two identical *wh*-phrases exchange their semantic roles. Semantic markedness surely plays a role in interpretive optimisation. An appropriate constraint should be ranked highly enough to rule out the interpretation in (67) early. This avoids the blocking of (58c) by a candidate like *c*₂ in (66). Candidate *c*₁ now becomes optimal.

The analysis presented here requires a fully parallel conception of the bidirectional architecture. The candidate set of feedback optimisation is not restricted to winners of the first optimisation. Herein the proposed model differs from serial bidirectional models like that of Wilson (2001) or Zeevat (2001). We also see that the bidirectional model is really a model of *syntactic* bidirectionality. Interpretation, here understood as interpretation of a surface form, is syntactic parsing and semantic interpretation at the same time.

7 Conclusion

I explored in this paper how a correspondence based conception of OT syntax derives classical minimality effects. I showed that the idea of structure preservation can be reconstructed in terms of correspondence between different representations. In standard “GB” (Chomsky, 1981) and minimalism (Chomsky, 1995), the configuration that underlies minimality effects among NPs is a projection according to the thematic hierarchy. Elements are inserted into their Θ -positions. This yields the asymmetric c-command relations between arguments that induce minimality effects in case- and operator-movement. This analysis is reconstructed here in terms of $M \Rightarrow S$ and $M \Rightarrow P$ mapping. The semantic relations encoded at M have to be translated into syntactic relations in terms of an iconic mapping that translates, for instance, relative scope and argument hierarchy into asymmetric c-command and precedence.

I then showed how this model extends to cases of word order freezing. These are

hard to account for in unidirectional models like minimalism and standard OT. The reason is that freezing is dependent on the surface form. It occurs if a marked syntactic structure is homophonous to a less marked one. Unidirectional grammars model syntax as a mapping from S to P, as a feeding relation. They do not take into account whether the reverse mapping, from P to S, also holds. Word order freezing is a case where this reverse mapping fails. As this is obviously crucial for grammaticality, the definition of grammaticality has to take into account both directions.

This has been done with a bidirectional version of OT syntax here. Two different kinds of minimality effects, superiority and word order freezing, are treated in an analogous fashion. In both cases, the constraint $M \Rightarrow S(\text{Arg})$ plays a crucial role. It takes effect in first optimisation in the case of superiority, and in feedback optimisation in the case of freezing.

The bidirectional perspective brings the surface form into the centre of attention. All underlying information needs to be encoded in terms of the surface form, and this must be done in such a way that it is recoverable. This defines grammaticality. The conditions for these translation processes can be formulated as OT correspondence constraints in the way demonstrated in this paper.

While this surface orientation is a new perspective in OT syntax in the radical way proposed here, other features of the model are very much inspired by previous work: LFG-OT makes extensive use of correspondence (Bresnan, 2000; Kuhn, 2001). The constraint $M \Rightarrow S(\text{Wh})$ expresses more or less the same as Grimshaw's (1997) 'OpSpec'. Likewise, $S \Rightarrow P(\text{hc})$ is reminiscent of Grimshaw's 'HeadLeft', though the two do not express exactly the same. The idea of specifying syntactic features in the input is present in (Baković and Keer, 2001; Legendre et al., 1998).

References

- Aissen, Judith 1987 *Tzotzil Clause Structure*. Dordrecht: Reidel.
- Aissen, Judith 1992 *Topic and Focus in Mayan*. *Language*, 68:43–80.
- Aissen, Judith 1996 *Pied-Piping, Abstract Agreement, and Functional Projections in Tzotzil*. *Natural Language and Linguistic Theory*, 14:447–491.
- Baković, Eric and Keer, Edward 2001 *Optionality and Ineffability*. In: Legendre et al. (2001).
- Barbosa, Pilar, Fox, Danny, Hagstrom, Paul, McGinnis, Martha, and Pesetsky, David (eds.) 1998 *Is the best good enough? Optimality and competition in syntax*. Cambridge, Massachusetts: MIT Press.
- Bresnan, Joan 2000 *Optimal Syntax*. In: Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, Syntax and Acquisition*, pp. 334–385. Oxford: Oxford University Press.
- Büiring, Daniel 2001 *Let's Phrase it. Focus, Word Order, and Prosodic Phrasing in German Double Object Constructions*. In: Gereon Müller and Wolfgang Sternefeld (eds.), *Competition in Syntax*, pp. 69–106. Berlin: Mouton de Gruyter.
- Choi, Hye-Won 1996 *Optimizing Structure in Context: Scrambling and Information Structure*. Ph.D. thesis, Stanford University.
- Chomsky, Noam 1981 *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, Noam 1986 *Barriers*. Cambridge, Massachusetts: The MIT Press.

- Chomsky, Noam 1995 *The Minimalist Program*. Cambridge, Massachusetts: MIT Press.
- Chomsky, Noam 2001 *Beyond Explanatory Adequacy*. Manuscript, MIT.
- Costa, Joao 2001 *The Emergence of Unmarked Word Order*. In: Legendre et al. (2001), pp. 171–203.
- Fanselow, Gisbert 1991 *Minimale Syntax*. *Groninger Arbeiten zur Germanistischen Linguistik*, 32.
- Greenberg, Joseph H. 1963 *Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements*. In: Joseph H. Greenberg (ed.), *Universals of Language*, pp. 58–90. Cambridge, Mass.: M.I.T. Press.
- Grimshaw, Jane, 1990 *Argument Structure*. Cambridge, MA: The MIT Press.
- Grimshaw, Jane 1991 *Extended Projection*. Manuscript, Brandeis University.
- Grimshaw, Jane 1997 *Projection, Heads and Optimality*. *Linguistic Inquiry*, 28:373–422.
- Haider, Hubert 1996 *Towards a Superior Account of Superiority*. In: Uli Lutz and Gereon Müller (eds.), *Papers on Wh-Scope Marking*, pp. 317–329. Arbeitspapiere des SFB 340, Bericht Nr. 76, Universität Tübingen.
- Haider, Hubert 2000 *Towards a Superior Account of Superiority*. In: Uli Lutz, Gereon Müller, and Arnim von Stechow (eds.), *Wh-Scope Marking*, *Linguistik aktuell/Linguistics Today*, pp. 231–248. Amsterdam, Philadelphia: John Benjamins.
- Hayes, Bruce, Tesar, Bruce, and Zuraw, Kie, 2002 *OTSoft*. *package*. <http://www.linguistics.ucla.edu/people/hayes/otsoft/>.
- Heck, Fabian and Müller, Gereon, 2000 *Successive Cyclicity, Long-Distance Superiority, and Local Optimization*. *Proceedings of WCCFL*, 19:218–231.
- Jackendoff, Ray, 1990 *Semantic Structures*. Cambridge, Massachusetts: MIT Press.
- Kayne, Richard, 1994 *The Antisymmetry of Syntax*. Cambridge, Massachusetts: MIT Press.
- Kuhn, Jonas, 2001 *Generation and Parsing in Optimality Theoretic Syntax: Issues in the Formalization of OT-LFG*. In: Sells (2001), pp. 313–366.
- Larson, Richard, 1988 *On the Double Object Construction*. *Linguistic Inquiry*, 19:335–391.
- Lee, Hanjung, 2001a *Markedness and Word Order Freezing*. In: Sells (2001), pp. 63–128.
- Lee, Hanjung, 2001b *Optimization in Argument Expression and Interpretation: A Unified Approach*. Ph.D. thesis, Stanford University.
- Legendre, Géraldine 2000 *Morphological and Prosodic Alignment of Bulgarian Clitics*. In: Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer (eds.), *Optimality Theory: Phonology, Syntax and Acquisition*, pp. 422–462. Oxford: Oxford University Press.
- Legendre, Géraldine, Grimshaw, Jane, and Vikner, Sten (eds.), 2001 *Optimality Theoretic Syntax*. Cambridge/Mass.: MIT Press.

- Legendre, Géraldine, Smolensky, Paul, and Wilson, Colin, 1998 *When is less more? Faithfulness and Minimal Links in WH-Chains*. In: Barbosa et al. (1998), pp. 249–289.
- McCarthy, John and Prince, Alan, 1995 *Faithfulness and Reduplicative Identity*. In: Jill Beckman, Laura Walsh-Dickie, and Suzanne Urbanczyk (eds.), *Papers in Optimality Theory*, vol. 18, pp. 249–384. Amherst, Massachusetts: UMass Occasional Papers in Linguistics.
- McCarthy, John J., 2002 *A Thematic Guide to Optimality Theory*. Research Surveys in Linguistics. Cambridge, UK: Cambridge University Press.
- Mondloch, James, 1978 *Disambiguating subjects and objects in Quiché*. *Journal of Mayan Linguistics*, 1(1):3–19.
- Müller, Gereon, 2001 *Order Preservation, Parallel Movement, and the Emergence of the Unmarked*. In: Legendre et al. (2001), pp. 279–313.
- Müller, Gereon, 2002 *Free Word Order, Morphological Case, and Sympathy Theory*. In: Gisbert Fanselow and Caroline Féry (eds.), *Resolving Conflicts in Grammars*, Linguistische Berichte Sonderheft 11, pp. 9-48, Hamburg: Helmut Buske Verlag.
- Pesetsky, David, 1997 *Optimality Theory and Syntax: Movement and Pronunciation*. In: Diana Archangeli and D. Terence Langendoen (eds.), *Optimality Theory. An Overview*, chap. 5. Malden, Mass. and Oxford: Blackwell.
- Pesetsky, David, 1998 *Some Optimality Principles of Sentence Pronunciation*. In: Barbosa et al. (1998), pp. 337–383.
- Prince, Ellen F., 1999 *How not to mark topics: ‘Topicalization’ in English and Yiddish*. Texas Linguistics Forum 1999, chapter 8.
- Rackowski, Andrea and Travis, Lisa, 2000 *V-initial languages: X or XP-movement and Adverbial Placement*. In: Andrew Carnie and Eithne Guilfoyle (eds.), *The Syntax of Verb Initial Languages*, pp. 117–142. New York: Oxford University Press.
- Rizzi, Luigi, 1990 *Relativized Minimality*. Cambridge, Massachusetts: MIT Press.
- Schmid, Tanja and Vogel, Ralf, 2004 ‘Dialectal Variation in German 3-Verb Clusters. A surface oriented OT account.’ *Journal of Comparative Germanic Linguistics*, 7:235-274.
- Sells, Peter (ed.), 2001 *Formal and Empirical Issues in Optimality-Theoretic Syntax*. Stanford, California: CSLI Publications.
- Smolensky, Paul, 1996 *On the Comprehension/Production Dilemma in Child Language*. *Linguistic Inquiry*, 27:720–731.
- Stepanov, Arthur, 2001 *Late Adjunction and Minimalist Phrase Structure*. *Syntax*, 4(2):94–125.
- Truckenbrodt, Hubert, 1999 *On the Relation between Syntactic Phrases and Phonological Phrases*. *Linguistic Inquiry*, 30:219–255.
- Uriagareka, Juan, 1999 *Multiple Spell-Out*. In: Samuel David Epstein and Norbet Hornstein (eds.), *Working Minimalism*, pp. 251–282. Cambridge, Mass.: MIT Press.
- Vogel, Ralf, 2004 *Remarks on the Architecture of Optimality Theoretic Syntax*. In: Reinhard Blutner and Henk Zeevat (eds.), *Optimality Theory and Pragmatics*, pp. 211-227. Houndmills, Basingstoke, Hampshire, England: Palgrave MacMillan.

- Vogel, Ralf, 2002 *Free Relative Constructions in OT Syntax*. In: Gisbert Fanselow and Caroline Féry (eds.), *Resolving Conflicts in Grammar. Optimality Theory in Syntax, Morphology, and Phonology*, Linguistische Berichte Sonderheft 11, pp. 119-163. Hamburg: Helmut Buske Verlag.
- Vogel, Ralf and Steinbach, Markus, 1998 *The Dative - an Oblique Case*. *Linguistische Berichte*, 173:65–90.
- Williams, Edwin, 2003 *Representation Theory*. Cambridge, Massachusetts: The MIT Press.
- Wilson, Colin, 2001 *Bidirectional Optimization and the Theory of Anaphora*. In: Legendre et al. (2001).
- Zeevat, Henk, 2001 *The asymmetry of Optimality Theoretic syntax and semantics*. *Journal of Semantics*, 17:243–262.
- Zepter, Alexandra, 2003 *Phrase Structure Directionality: Having a few Choices*. Ph.D. thesis, Rutgers University, New Brunswick, NJ.