Remarks on the Architecture of OT Syntax Grammars

Ralf Vogel

This paper argues for a particular architecture of OT syntax. This architecture has three core features: i) it is bidirectional, the usual production-oriented optimisation (called 'first optimisation' here) is accompanied by a second step that checks the recoverability of an underlying form; ii) this underlying form already contains a full-fledged syntactic specification; iii) especially the procedure checking for recoverability makes crucial use of semantic and pragmatic factors.

The first section motivates the basic architecture. The second section shows with two examples, how contextual factors are integrated. The third section examines its implications for learning theory, and the fourth section concludes with a broader discussion of the advantages and disadvantages of the proposed model.

1 Syntax in Optimality Theory – A Proposal

An OT system maps an input to an output according to a system of hierarchically ordered criteria. Such systems can be developed for the modelling of many different things, not only linguistic processes. A central question for the design of an OT system is the choice of the objects serving as input and output and their representational formats. OT systems that use the same objects for input and output have to be distinguished from those that use different ones.

In much of the work in OT phonology, input and output consist of the same elements. For example, the mapping from the input “[tag]” to the output “/tak/” in German describes the process of final devoicing by using strings of phonological segments in both input and output. In their discussion of syllabification, Prince & Smolensky (1993/2002) use output representations that contain the input representations and enrich them with syllable structure. Thus, the plural form for German “/tag/”, “[tag+ə]” is mapped onto “/ta.gə/”. Other tasks require syllable structure already in the input. One example is the description of loan word integration. Languages that avoid codas and complex onsets resyllabify loan words with such properties. Kenstowicz & Sohn (1998) show this for the Korean dialect of North Kyungsang, where, for example, the name “kris.to” is turned into “ku.ri.su.to”.

In OT syntax, a model that has often been used is that of a mapping from a *semantic* representation in the input to a *syntactic* representation in the output (Grimshaw, 1997). Here, input and output are radically different. The input-output mapping has the character of a *translation*.

But just as in the case of loan words shown above, it might also sometimes be useful to have the same types of representations, for example, if one wants to describe the typology of syntactic constructions: If language A lacks a particular construction C that occurs in language B, an OT model could show that C would be mapped onto a different construction D if it was in the input in language A.[1]

One example in case is the typology of free relative constructions as modeled in Vogel (2001, to appear):

(1) German free relative and correlative construction:

a. Wer einmal lügt, lügt auch zweimal
who-NOM once lies lies also twice

b. Wer einmal lügt, der lügt auch zweimal
who-NOM once lies that-one-NOM lies also twice

Free relative constructions (FR) as in (1-a) are marked compared to correlative constructions (CR) as in (1-b): Languages that have FRs also have CRs, but there are languages with CRs that lack FRs. Also, languages with FRs differ in the contexts which allow for this construction – contexts which allow for free relatives also allow for correlatives, but there are contexts allowing for correlatives that do not allow for free relatives. For example, in German, a FR is out, if it would imply the suppression of oblique case (in the following example, dative):

(2) Wer einmal lügt, *(dem) glaubt man nicht
who-NOM once lies the-one-DAT believes one not

The solution I proposed in the works cited above is an OT system where the syntactic structure (FR or CR) is specified in the input, and where FRs and CRs compete in the output. In cases like (2), a FR in the input is neutralised to a CR in the output. A CR in the input, however, is always mapped onto a CR in the output.

Another source of the plurality of architectures is the fact that OT syntacticians come from different frameworks. OT syntax work has been done within Government and Binding Theory, Minimalism, Lexical Functional Grammar, Functional Grammar and possibly even more frameworks (representative examples can be found in the collections by Legendre et al., 2001; Dekkers et al., 2000; Sells, 2001). These frameworks essentially differ in the character, number and formats of representations that they use.

My impression of current OT syntax work is, nevertheless, that OT systems developed within the different frameworks can usually be translated in a straightforward way without any damage to the systems themselves. The explanatory value of an OT model is usually independent of the representational 'language' that is used. Very often, OT constraints are defined in a quite informal way. This makes the translation from one framework into the other quite easy. In fact, the choice of framework seems to become a minor issue.

This is an expected outcome insofar as the explanatory burden is shifted from assumed properties of representations to constraint interaction. The question of what is the appropriate representation for a particular syntactic construction has less 'weight' within the theory. But this also means that representations can be simplified if one uses OT in explaining syntactic phenomena.

On the other hand, as long as OT syntax work looks so diverse, and is not formulated independently of non-OT frameworks, OT in syntax looks more like a *method* adapted *within* different 'traditional' frameworks than like a framework in its own right. What might be achievable in approaching the latter aim, is the development of a kind of 'meta-language' for syntactic representations.

Which representations does an OT syntax system actually need? I want to follow Jackendoff (1997) who summarises the traditional point of view of what grammars are doing: he claims that there are three representations, a semantic, a syntactic and a phonological representation, and it is their *correspondence* that is modeled by a theory of grammar. Let us use the symbols **M** (for 'meaning'), **S** (syntax) and **P** (phonology) for these representations. The syntactic frameworks mentioned above differ in their assumptions about **S**, its complexity and format, and in how much of **P** and **M** enters the considerations about **S** and its role in grammar.

In Grimshaw (1997), the input can roughly be identified with **M**, it contains argument structural information. Information structural specifications are included in later work of Grimshaw (see Grimshaw & Samek-Lodovici, 1998). The output candidates come close to what is called ‘S-structure’ in Government-Binding Theory (Chomsky, 1981). S-structure covers some aspects of **P**, namely, morphology and linear order. But prosodic and metrical structure are not represented at all.

Pesetsky (1997, 1998) models a particular aspect of minimalist grammars (Chomsky, 1995) in an OT fashion, namely, the mapping from LF (‘Logical Form’, an abstract syntactic representation) to PF (‘Phonetic Form’), which can be rephrased as the correspondence between **S** and **P**. The empirical coverage of Pesetsky’s work is rather small, touching only on aspects of the overt realisation of lexical elements, but the model has more general implications. Truckenbrodt (1999) models the correspondence of syntactic and prosodic phrases, Büring (2001), Samek-Lodovici (2002) and Schmid & Vogel (submitted) use similar systems in their discussion of the relation between focus and word order.[2] While Pesetsky’s OT model is a ‘partial’ grammar in the sense that it models a mapping from **S** to **P**, without using **M**, the mentioned works on focus use at least the information structural aspects of **M**. The approaches differ in whether **S** is part of the input (e.g., Pesetsky), or part of the output (e.g., Grimshaw). If **M** is the only input representation, then the output is a pair [**S**,**P**], but if **P** is the one and only output representation, then the input must be a pair [**M**,**S**].

These considerations illustrate a common assumption about the role of syntax as *mediating* between ‘meaning’ and ‘sound’. One way of modelling this could be a serialisation of two optimisations, one where **M** is mapped onto **S**, and a second step, where the winning **S** is mapped onto **P**. This would imply that there is no direct correspondence relation between **M** and **P**. But the works on focus mentioned above make crucial use of constraints reflecting the correspondence of **M** and **P** – it is uncontroversial that prosodic structure directly reflects information structure. The picture that we get looks more like a triangle: **M** is connected with both **S** and **P**, as are **S** and **P**.

In my work on free relative constructions discussed above (Vogel, 2001, to appear) I show the need for having **S** in *both* input and output. The mediating function of **S** is reflected by this double occurrence. The main motivation for this structure, however, is the need to implement a basis for optionality and ineffability of syntactic constructions. In the case of FRs and CRs introduced above, it is obvious that the two constructions stand in a markedness relation: FRs are more marked than CRs, and CRs can always be inserted for FRs, but not always vice versa. The two constructions only differ formally. Universally, the set of languages that have FRs is a proper subset of languages with CRs, and within a particular language, the set of contexts that allow for FRs is a subset of those that allow for CRs.[3]

For a marked structure to survive the competition against the unmarked one, it must be given some advantage, which is usually done by specifying it in the input. Faithfulness constraints ensure that the marked structure wins, as long as the markedness constraints that this structure violates are ranked lower than the faithfulness constraints that are violated by a less marked candidate.

The model that I propose for OT syntax combines two basic issues: having a way of accounting for optionality and ineffability in a standard OT fashion, and implementing the mediating function of syntax. In sum, the structure of input and output (candidates) is the following (the two occurrences of **S** are distinguished by subscripts):

(3) Input and output representations in OT syntax, (see Vogel, to appear):

Input:	$\mathbf{S}_I$, **M**
Output:	$\mathbf{S}_O$, **P**

The models discussed thus far share the property of being *uni-directional* models. Recent work has suggested that for some purposes a *bidirectional* perspective is necessary. Especially Wilson (2001), Kuhn (2001) and Lee (2001a,b) have to be mentioned here. 'Bidirectional' means here that besides an optimisation from meaning to form, OT syntax needs a second optimisation from form to meaning. Applications of this idea are still quite rare. Wilson (2001) uses a serial model where optimisation from meaning to form restricts the candidate set for the second, syntactic optimisation. The model that I argue for in this paper, uses interpretive optimisation as a 'post-filter' mechanism. This idea also has predecessors.

Pesetsky (1997, 1998) introduced a constraint that he called RECOVERABILITY, which requires semantically relevant material in **S** to be 'visible' at **P**.[4]

But recoverability can only be checked in a process that reverses the direction of optimisation: the original output serves as input, and the original input should be the optimal output of the former's optimisation. If this is the case, then recoverability is proven. Lee (2001b; 2001a, see also Beaver & Lee, this volume) shows that not only semantic aspects are subject to the recoverability condition, but also syntactic ones. An underlying object-subject order might not be recoverable from **P**, if subject-object order is the unmarked case in a language, and if there are no morphological or other hints that signal the underlying marked order – a classical case of neutralisation. The following German example combines the two aspects of recoverability:

(4) Zwei Professoren haben drei Studenten
two professors have three students

The default interpretation for a clause like (4) is that it has subject-object order and a quantifier scope that follows the linear order of the quantifiers. However, the two NPs are ambiguous for nominative and accusative, and object-subject order is not ungrammatical in principle. Likewise, scope reversal would be possible under other circumstances, or with the help of contextual factors.[5] Thus, an input that is specified for object-subject order and inverse scope relations should be able to survive. That the structure in (4) does not have this interpretation in the default case results from a second step of optimisation. In this second step, we are looking for the optimal underlying structure of a given surface form. Here the input is the winning **P** of the initial optimisation process and we look for the optimal underlying pair [**S**, **M**]. I call this second step *feedback optimisation* (see Vogel, 2002). This grammar has the following structure:

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

First optimisation:	Input:	$\mathbf{S}_I$, **M**
	Output:	$\mathbf{S}_O$, **P**
Feedback optimisation:	Input:	**P**
	Output:	$\mathbf{S}_I$, **M**

The model emphasises the role of **P** as the *ultimate* representation in terms of which all underlying information, both semantic and syntactic, has to be encoded. **P** includes *all* aspects of the 'surface form', in particular, it is also the only representation that encodes linear order. This is a common assumption in contemporary generative syntax (cf., e.g., the work based on Kayne, 1994). In these models, the abstract syntactic representation only encodes dominance and relations derived from this, like constituency and c-command, furthermore, it contains the abstract features of lexical items, and syntactic categories.

(6) Assumed Representations and what they represent:

M: argument structure, scope relations, information structure etc.

S: constituency, abstract features, syntactic categories etc.

P: linear order, overt morphology, prosodic structure etc.

There are many ‘natural’ ways of *encoding* relations within these representations. For example, the semantic relations quantifier scope and argument structure are usually translated into (asymmetric) c-command at **S** and precedence at **P**. Likewise, predication is encoded into sisterhood at **S** and adjacency at **P**. Assuming correspondence constraints that formulate these ‘default translations’ is straightforward. We will turn to some examples in the next section.

It is crucial that the same constraint hierarchy is used in both optimisation steps. The recoverability condition is implemented into this model as a condition on grammaticality:

(7) **Grammaticality:**

A triple [$\mathbf{M}_i$,$\mathbf{S}_i$,$\mathbf{P}_i$] is grammatical, if and only if the input [$\mathbf{M}_i$,$\mathbf{S}_i$] yields [$\mathbf{S}_i$,$\mathbf{P}_i$] in first optimisation, and the input [$\mathbf{P}_i$] yields [$\mathbf{M}_i$,$\mathbf{S}_i$] in feedback optimisation.

Ungrammaticality may arise in both optimisation steps. An $\mathbf{S}_I$ might be mapped onto a different $\mathbf{S}_O$ in first optimisation – ungrammaticality of a particular syntactic structure; or $\mathbf{S}_I$ wins the first optimisation, but loses the feedback optimisation of its winning **P** – a case of unrecoverability under particular circumstances, usually connected to indeterminacies given in the surface form. The next section discusses example applications of this model. It will also show that the model may not be viewed as ‘encapsulated’. Especially markedness constraints on **M** have to make crucial use of information provided by context and world knowledge.

2 Two Examples

2.1 Word Order Freezing

Let us first consider a simple case of word order freezing in German:

(8) a. Den Hans liebt Maria
the-ACC H. loves M.
‘As for Hans, Maria loves him’

b. Hans liebt Maria
H. loves M.
‘Hans loves Maria’

Both ‘*Hans*’ and ‘*Maria*’ are ambiguous for nominative and accusative case in (8-b). Without contextual disambiguation, (8-b) cannot be interpreted like (8-a). The unmarked case is subject-object order. A marked order requires disambiguation, in (8-a) the determiner marks the initial NP as accusative. The fronting of ‘*den Hans*’ reflects the topic status of that NP.

I will now reconstruct this case using the following constraints on the *correspondence* of **M** and **S**:

(9) Constraints on M⇨S mapping:

(elements of **M** are called, ‘m_n’, elements of **S**, ‘s_n’, and elements of **P**, ‘p_n’; identical indices indicate correspondence of elements, e.g., m_1 corresponds to s_1)

a. Arg⇨**S**: If an argument m_1 is higher than another argument m_2 at **M**, then s_1 asymmetrically c-commands s_2 at **S**.

b. Inf⇨**S**: If m_1 is [+topic] and m_2 is [–topic] at **M**, then s_1 asymmetrically c-commands s_2 at **S**.

These two constraints conflict in the case of (8-a), where the lower argument, the object, is topic. That this clause is grammatical, shows that the order of the two constraints in German must be:

(10) Inf⇨**S** $\gg$ Arg⇨**S**

If the ranking was the other way around, then such a structure could not survive the first optimisation: it would lose against a subject-initial structure. In feedback optimisation, we have **P** in the input and search for the optimal underlying form, a pair [**M**,**S**]. Here, the only difference between (8-a) and (8-b) is important: the determiner, which signals the case of the initial NP. The correct 'translation' of the surface morphology into underlying abstract syntactic features is evaluated by a constraint on **S**⇨**P** correspondence. The bare noun '*Hans*' fits both nominative and accusative, so neither of these two 'interpretations' would violate **S**⇨**P** for (8-b). Likewise, the initial NP '*Hans*' can be interpreted as topic, independent of its grammatical function, Inf⇨**S** cannot be decisive either, and so finally Arg⇨**S** makes the decision favouring a subject-initial structure:

(11) Feedback optimisation for (8-b):

Hans liebt Maria	S⇨P	INF⇨S	ARG⇨S
OVS, O=topic			∗!
☞SVO, S=topic			

But in the case of '*den Hans*' in (8-a), **S⇨P** is violated by the candidate that interprets this NP as nominative instead of accusative, and so the OVS candidate is the winner:

(12) Feedback optimisation for (8-a):

Den Hans liebt Maria	S⇨P	INF⇨S	ARG⇨S
☞OVS, O=topic			∗
SVO, S=topic	∗!		

S⇨P is an interesting constraint, because its classification as faithfulness or markedness constraint is different in the two optimisation steps. Markedness constraints only evaluate properties of candidates irrespective of the input. In this respect, **S⇨P** behaves like a markedness constraint in first optimisation.[6] In feedback optimisation, **P** is in the input and **S** in the output. **S⇨P** now acts as a faithfulness constraint.

Another important aspect of this perspective on grammaticality is its context dependency. The effects of word order freezing can be overcome. In the context of a question like (13), the preference for the interpretation of (8-b) is clearly object-subject order.

(13) Wen liebt Maria?
who-ACC loves M.

Let us assume that the context, a discourse representation of whatever format one prefers, is present and accessible for constraint evaluation. We can then formulate a constraint like (14):

(14) **M**fits**C**: **M** is compatible with the context **C**

This constraint is a markedness constraint on possible interpretations. It favours interpretations that fit into a given context over others that do not fit. It only plays a role in feedback optimisation, as only here **M** is part of the candidates and therefore subject to evaluation. The constraint plays the same role as **S**⇨**P** in the example we had before, in preserving the marked underlying OVS order:

(15) Feedback optimisation for (8-b) in the context (13):

Hans liebt Maria	MfitsC	INF⇨S	ARG⇨S
☞OVS, O=topic			*
SVO, S=topic	*!		

Likewise, such a preference can be triggered by world knowledge, as in (16), where only the second NP can meaningfully be interpreted as having the experiencer role of *love*:

(16) Fussball liebt Maria
football-NOM/ACC loves M.-NOM/ACC
'Football, Maria loves'

Let us assume that another markedness constraint on **M** plays the decisive role here, which is similar to **M**fits**C**. It can roughly be formulated as '**M** fits the world'.

This model of grammaticality assumes that we use all resources we can in order to recover underlying structure. At least the second step of optimisation is nonencapsulated, and in this respect the model differs from the traditional generative grammarian point of view.

This is not a model of semantic interpretation, it is a model of grammaticality. But it makes use of semantic and pragmatic factors, because it assumes that these factors are crucial for grammaticality to a certain extent. Grammars may differ in the role pragmatics plays for grammaticality. For Russian, which allows for object-subject orders in principle, it has been claimed that a clause like (16) is ungrammatical under case ambiguity in "non-emotive speech" (cf. Bloom, 1999). World knowledge obviously does not help in escaping word order freezing in Russian, which would mean that the respective constraint is ranked lower than ARG⇨**S**.

2.2 Superiority and Discourse-Linking

The paradigm in (17) displays a well-studied contrast in the syntax of English multiple questions:

(17) a. *What did who do?

b. What did which student do?

This contrast has been discussed in detail by Pesetsky (1987). His explanation for the difference between (17-a) and (17-b) is that (17-b) is grammatical, because the *which* NP is what he called 'discourse-linked' (d-linked): it refers to a set of individuals that has already been introduced in the preceding discourse. This, we infer from this argument, does not hold of *who* in (17-a). But Bolinger (1978) already showed that the empirical generalisation about (17-a) is also not as straightforward as people often think. He gives the example in (18) to show that this clause can be acceptable in a suitable context (capital letters indicate main stress):

(18) I know what just about everybody was ASKED to do, but what did who (actually) DO?

This example strengthens Pesetsky's point: here, *who* refers to individuals that have already been introduced into the discourse, and the clause is acceptable. The scenario that I want to reconstruct in this section has the following features:

- there are two forms, *who* and *which*
  - *who* is interpreted as non-d-linked by default, but can be interpreted as d-linked given the right context
  - *which* is interpreted as d-linked
- both elements are individual lexical items and as such can be part of the input
- the two elements are related on a markedness scale: *which* is more marked than *who*

This case is an example of 'partial blocking': *who* could be interpreted as d-linked, but the mere existence of *which* usually blocks it. Under particular conditions, however, this blocking can be overcome. *who* is assumed to be the unmarked form, because it goes along with non-d-linking, which seems to be the unmarked interpretation, though it is not the only one possible. *Which* can only be interpreted as d-linked. So we have two markedness scales:[7]

(19) a. *who*, *what* $\ldots <$ *which* NP

b. $-$d-linked $<$ $+$d-linked

These two scales can be used for the generation of constraints with the method of 'harmonic alignment', developed by Prince & Smolensky (1993/2002). In a first step, we build two sub-hierarchies of constraints, one for each form (‘dl’ is an abbreviation for ‘d-linked’):

(20) a. **who*/+dl $\gg$ **who*/$-$dl

b. **which*/$-$dl $\gg$ **which*/+dl

The two rankings in (20) are universally fixed, but their interaction is free.[8] Suppose that the ranking in English is the following:

(21) **which*/$-$dl $\gg$ **who*/+dl $\gg$ **who*/$-$dl $\gg$ **which*/+dl

(21) states, for instance, that the most marked case is the one where *which* is interpreted as non-d-linked. This is the only case that is not attested in English, as far as I can see. I assume that, although *who* and *which* are already specified in $\mathbf{S}_I$, they nevertheless compete in candidate sets.

The non-occurrence of non-d-linked *which* can be prohibited with a constraint on input preservation in **S**, $\mathbf{S}_I$⇨$\mathbf{S}_O$. It is ranked below **which*/$-$dl:

(22) **which*/$-$dl $\gg$ $\mathbf{S}_I$⇨$\mathbf{S}_O$ $\gg$ **who*/+dl $\gg$ **who*/$-$dl $\gg$ **which*/+dl

The predictions of this system are easy to detect: a [–dl] *which* input yields *who* as output. In all other cases, the output form is the one given in the input:

(23) First optimisation:

Input: *which*,+dl $\longrightarrow$ *which*

Input: *which*,$-$dl $\longrightarrow$ *who*

Input: *who*,+dl $\longrightarrow$ *who*

Input: *who*,$-$dl $\longrightarrow$ *who*

In feedback optimisation, we take the form we obtained as input and look for the best interpretation, i.e., either d-linked or non-d-linked. As there is no faithfulness involved here, it is clear that *who* yields [$-$dl], and *which* yields [$+$dl]:

(24) Feedback optimisation:

who $\longrightarrow$ $-$dl

which $\longrightarrow$ $+$dl

Our model of grammaticality combines the two perspectives, and treats as grammatical only those [input, output] pairs where the input is recoverable from the output. Only two of the four cases in (23) have this property, namely, (25-a,d):

(25) First plus feedback optimisation:

a. Input: *which*,$+$dl $\longrightarrow$ *which* $\longrightarrow$ $+$dl

b. Input: *which*,$-$dl $\longrightarrow$ *who* $\longrightarrow$ $-$dl

c. Input: *who*,$+$dl $\longrightarrow$ *who* $\longrightarrow$ $-$dl

d. Input: *who*,$-$dl $\longrightarrow$ *who* $\longrightarrow$ $-$dl

This system derives the default interpretations that we observed for the *wh*-phrases under examination. One reading is missing, namely, the contextually forced [$+$dl] interpretation for *who*, as exemplified in (18). It will be preserved, if contextual information is taken into account. To include this, we introduced the general constraint ‘MfitsC’ in the previous section which may also be used here. It is ranked on a par with $\mathbf{S}_I$⇨$\mathbf{S}_O$:

(26) **which*/$-$dl $\gg$ MfitsC $\mathbf{S}_I$⇨$\mathbf{S}_O$ $\gg$ **who*/$+$dl $\gg$ **who*/$-$dl $\gg$ **which*/$+$dl

Feedback optimisation within the right context gives *who* the chance to be interpreted as [$+$dl] (27-a):

(27) Feedback optimisation, including context:

a. *who*, context: $+$dl $\longrightarrow$ $+$dl

b. *who*, context: $-$dl $\longrightarrow$ $-$dl

c. *which*, context: $+$dl $\longrightarrow$ $+$dl

d. *which*, context: $-$dl $\longrightarrow$ $+$dl

The discussion in this subsection demonstrates that harmonic alignment can implement the 'division of pragmatic labour' (Horn, 1984), the observation that unmarked forms tend to be used for unmarked situations and marked forms for marked situations. Harmonic alignment can be an alternative to 'weak bidirectional systems' (see also Beaver & Lee, this volume) in the sense of Blutner (2001). The most important effect of a weak bidirectional system – modeling of the division of pragmatic labour – can be implemented within a strong bidirectional system like the one developed in this article. One prerequisite for this possibility is that the forms and interpretations in question can sensefully be compared in terms of a single parameter of markedness. For the standard example discussed in Blutner (2001), this is the case. The example is:

(28) a. Black Bart killed the sheriff

b. Black Bart caused the sheriff to die

The two clauses differ in meaning: (28-b) has an interpretation where the causation is much more indirect than in the case of (28-a). The two markedness scales that we can use for harmonic alignment here are:

(29) a. $[_{VP}$ V$] < [_{VP}$ V $[_{VP}$ V $]]$

('simple VP is less marked than complex VP')

b. direct causation $<$ indirect causation

Using these scales, we can construct constraints as exemplified above, get a fixed ranking in the desired way and derive the wanted effect.

3 Bidirectional OT Syntax and Learning Theory

The bidirectional model of OT syntax that has been developed in the previous sections is reminiscent of models that have been explored in OT learning theory. Tesar & Smolensky (2000) describe the learning of an OT system as the iterated application of a three-step process in the following way:

(30) The Constraint Demotion/Robust Interpretive Parsing (CD/RIP) OT learning procedure, (after Tesar & Smolensky, 2000, 62):

Given an overt form **OF** and an (initially arbitrary) constraint ranking, **H**,

a. The learner assigns to **OF** a structural description $\mathbf{SD}_I$ including an underlying form **UF**.

b. The learner then applies production directed optimisation to **UF** and yields another structural description $\mathbf{SD}_P$.

c. If $\mathbf{SD}_P$ is identical to $\mathbf{SD}_I$, then **H** does not need adjustment.

d. If $\mathbf{SD}_I$ and $\mathbf{SD}_P$ differ, then an error has occurred, the learner needs to adjust **H**. She assumes $\mathbf{SD}_I$ to be correct and applies

e. Constraint demotion, with $\mathbf{SD}_I$ as winner and $\mathbf{SD}_P$ as loser: constraints that are violated (more often) by $\mathbf{SD}_I$ are reranked below constraints that are violated by $\mathbf{SD}_P$.

It needs to be shown that the OT syntax model proposed here fits into this general description of a learnable OT grammar. What I called ‘feedback optimisation’ can be identified as the initial step (30-a) in Tesar & Smolensky's (2000) learning procedure. **P** would then be the overt form, the current constraint ranking would be used to get an interpretation for that overt form, a pair [**S**,**M**]. However, the overt form in that model is a 'surface reflection', only the overt part of the winning candidate, and as such, it cannot be subject to constraint evaluation, unlike **P**.[9] Thus, the overt form cannot be **P** itself, but only its 'reflection'. **P** is part of the structural description of a clause, as well as **S** is.

The interpretation $\mathbf{SD}_I$ should then be identified with the triple [**M**,**S**,**P**]. It contains the underlying form **UF** = [**M**,**S**]. The second step in the algorithm applies production oriented optimisation, my 'first optimisation', to [**M**,**S**], yielding a structural description $\mathbf{SD}_P$ = [**S**,**P**]. Step (30-c) needs slight revision. $\mathbf{SD}_P$ cannot be identical to $\mathbf{SD}_I$, because the latter is a triple [**M**,**S**,**P**], while the former is a pair [**S**,**P**]. Hence, instead of the identity of $\mathbf{SD}_P$ and $\mathbf{SD}_I$, we have to check for the identity of the relevant parts of the two representations. This is in fact the only adjustment that would have to be made, and it appears rather harmless to me. Of course, the major underlying assumption of the whole approach is that the representations we are dealing with are quite complex objects. But this is fairly uncontroversial in the area of syntax.

Tesar & Smolensky (2000, 63) mention three scenarios where the algorithm fails. These are the following:

- **Selecting an interpretation that cannot possibly be optimal**. This can happen with 'weird' optimal forms which are highly marked. The learner nevertheless assigns an interpretation to it. But this interpretation will not survive the second optimisation process. This causes reranking, which then causes, in the next cycle, a new interpretation for the overt form, which again does not survive, again constraint demotion applies and might reestablish the ranking we had before, and the system might run into an endless cycle till it stops.

- **The optimal interpretation is harmonically bound**. A winning interpretation is found to lose under any ranking in the second step of optimisation. This situation is easy to handle: the learner can give up learning on the particular data. There is no ranking that would derive the current interpretation as winner. The grammar cannot be learned with the particular data at hand.

- **Endless alternation between different overt forms**. This is another kind of endless circle. Two different data require different rankings, and trigger these whenever they are processed.

As Tesar & Smolensky already discussed, these situations are rather special. The second problem should not pose particular difficulties as long as it only rarely occurs within the set of training data. The first and the third problem point to possible inconsistencies in a language or the given data. Especially the third case is one where usually alternatives to strict ranking are considered, like, for instance, constraint ties or parallel grammars. Each of these cases might as well occur in syntax learning. For successful learning, it is important that cases like these are rare among the training data.

One further problem could be the acquisition of underlying forms. It is especially problematic in morphology, i.e., in the acquisition of ‘irregular’ lexical items, which have to be acquired as whole *paradigms*, not as single elements, crucially because of allomorphic variation. However, for OT syntax it has usually been assumed that underlying forms are universal, therefore need not be learned. For **M**, this is quite clear. For **S**, this is a debatable assumption among syntacticians. The generative tradition assumes that abstract syntactic structures are universal: this includes the inventory of syntactic categories and features, as well as the mechanisms of their combination into larger units. At least one proposal has been put forward recently, Croft's (2001) 'Radical Construction Grammar', that assumes that syntactic constructions are language particular, and thus have to be learned, just like lexical elements have to be learned. This is something that the model proposed here might also be able to live with, as long as constructions can be shown to be as learnable as lexical items in general. This task is beyond the scope of this paper, however.

4 Conclusion

Beaver & Lee (this volume) discuss different OT architectures and compare how they are able to deal with a number of phenomena. The model for OT syntax developed here belongs to their category of 'strong bidirectional models'. Beaver & Lee show that models of this category can successfully deal with freezing, blocking, uninterpretability and ineffability, but that they also fail in dealing with optionality, ambiguity and partial blocking. The model that I developed here interestingly is more successful in each of these three cases. Section 2.2 showed how at least simple cases of partial blocking can be dealt with by using the method of harmonic alignment. In accounting for the optionality of forms, I formulated the need for a 'double occurrence' of syntactic specifications in both input and output. A marked form specified in the input is preserved in the output by highly ranked faithfulness constraints.

A more difficult case is the ambiguity of a single form. A very hard case that has not been discussed in this paper yet, is context-independent ambiguity. A potential example is (31):

(31) Welche Frau hat Hans gesehen?
which woman-NOM/ACC has H.-NOM/ACC seen?
'Which woman saw Hans?' OR 'Which woman did Hans see?'

Although German observes freezing with two ambiguous proper nouns, the structural ambiguity is preserved, if (only!) one of the two NPs is a *wh*-phrase. The way out of this problem that I proposed in (Vogel, 2002) is redefining the constraints on syntactic ordering such that they only apply to elements of the same syntactic type. Thus, a constraint like 'ARG⇨**S**' would not be violated by any interpretation of (31), because the two NPs are of different type. One possible way of accounting for ambiguity is thus ensuring that the constraints make no decision between two candidate interpretations, by defining the constraints accordingly.

In section 2, I showed how partial blocking in the case of word order freezing and simple *wh*-elements can be overcome by referring to properties of the context. The claim is that contextual factors can uncover the underlying ambiguity of an expression. A well-known example from phonology which has been discussed by Zeevat (2001) (see also Beaver & Lee, this volume), results from the phenomenon of final devoicing in languages like German and Dutch. In Dutch, the phonetic string [rʌt] is ambiguous for the underlying forms /rʌd/ ('wheel') and /rʌt/ ('rat'). However, in 'real life' the two interpretations can usually be distinguished by quite easily contextual means. Once this context dependency is reflected in a grammar, in the form of constraints like 'MfitsC', there is a way to derive and predict the possibility of two or more interpretations of an expression.

I hope to have shown that such a reflection of pragmatic factors *within* an OT model of syntax is necessary and desirable. Syntax is much less encapsulated and 'autonomous' than generative grammar usually assumes. The discussion in section 2 suggests that the application of core syntactic constraints is restricted by pragmatic constraints. The picture of grammar that emerges from the considerations in this paper is that of a ‘total grammar’ where expressive and interpretive constraints collaborate and interact, and even syntax can only be understood from the perspective of this very global interaction. In turn, a pragmatic principle like the ‘division of pragmatic labour’ describes the mutual dependency of related meanings and forms. It receives a natural expression within bidirectional OT models.

Acknowledgments I thank Reinhard Blutner and Henk Zeevat, as well as the participants of the 2002 ”Optimality Theory and Pragmatics” workshop for helpful comments, suggestions and discussions. This work has been supported by grant Nr. FOR-375/1-A3 from the Deutsche Forschungsgemeinschaft (DFG) for the research project “Minimality in Optimality Theory” as part of the interdisciplinary research group “Conflicting Rules in Language and Cognition”, University of Potsdam.

Notes

[1]The first who proposed a model with such properties for OT syntax, were Baković & Keer (2001), as far as I know.

[2]It is not accidental that much of recent work in OT syntax is devoted to very ‘surfacy’ aspects of syntax. Radical surface orientation was the major change that OT induced in phonology. Proponents of this surface orientation, in addition to those researchers mentioned in the text, are Geraldine Legendre and Stephen Anderson (see, for example, Legendre (2001) and references cited there, and Anderson (2000)).

[3]This situation is fully parallel to typical cases of markedness in phonology. Con-

sider, for instance, the relation between voiced and voiceless obstruents. All languages that have voiced obstruents, also have voiceless obstruents, but there are languages with voiceless obstruents that lack voiced ones. Second, the contexts where voiced obstruents occur are very often more limited than those for voiceless ones. In German, for example, voiced obstruents only occur in the onset, but never in the coda of the syllable. Voiceless obstruents can occur in both positions. The syntactic example given in the text is only one among many others that could also be chosen: passive vs. active, object-subject orders against subject-object orders, complementiser-less subordinate clauses vs. complementiser-introduced clauses in English and German, etc.

[4]The definitions Pesetsky gives for the RECOVERABILITY constraint, are quite informal:

> "A syntactic unit with semantic content must be pronounced unless it has a sufficiently local antecedent." (Pesetsky, 1998, 342)
>
> "[. . .] This fact is accounted for by a principle called the Recoverability Condition – the idea being that the semantic content of elements that are not pronounced must be recoverable from local context. [. . .]" (Pesetsky, 1997, 154)

[5]One possible way of triggering scope inversion would be a question of the following form:

(i)	Wieviele	Studenten	sind	bei	zwei	Professoren?
	How many	students	are	at	two	professors?

[6]To be precise, **S**⇨**P** should be called $\mathbf{S}_O$⇨**P**. The role of $\mathbf{S}_I$ must be restricted to

constraints that belong to the $\mathbf{S}_I \Rightarrow \mathbf{S}_O$ family.

[7]The terms *who* and *which* as used in this 'universal' markedness scale should be understood as 'placeholders' for abstract universal functional categories.

[8]This means that if we have two fixed sub-rankings 'A1 $\gg$ A2' and 'B1 $\gg$ B2', there are six possible rankings:

(32) a. A1 $\gg$ A2 $\gg$B1 $\gg$ B2

b. B1 $\gg$ B2 $\gg$ A1 $\gg$ A2

c. A1 $\gg$ B1 $\gg$ A2 $\gg$ B2

d. B1 $\gg$ A1 $\gg$ B2 $\gg$ A2

e. A1 $\gg$ B1 $\gg$ B2 $\gg$ A2

f. B1 $\gg$ A1 $\gg$ A2 $\gg$ B2

[9]I thank Reinhard Blutner for making me aware of this problem.

References

Anderson, S. R. (2000). Towards an optimal account of second-position phenomena. In Dekkers et al. (2000), (pp. 302–333).

Archangeli, D. & Langendoen, D. T. (Eds.) (1997). *Optimality Theory. An Overview*. Malden, Mass. and Oxford: Blackwell.

Baković, E. & Keer, E. (2001). Optionality and ineffability. In Legendre et al. (2001), (pp. 97–112).

Barbosa, P., Fox, D., Hagstrom, P., McGinnis, M., & Pesetsky, D. (Eds.) (1998). *Is the best good enough? Optimality and competition in syntax*. Cambridge, Massachusetts: MIT Press.

Beaver, D. & Lee, H. (this volume). Input-output mismatches in OT.

Bloom, D. B. (1999). They're freezing in russia: A typology of word order "freezing" in Russia. Manuscript, Stanford University.

Blutner, R. (2001). Some aspects of optimality in natural language interpretation. *Journal of Semantics*, *17*, 189–216.

Bolinger, D. D. (1978). Asking more than one thing at a time. In H. Hiż (Ed.), *Questions*, (pp. 107–150). Dordrecht: D. Reidel.

Büring, D. (2001). Lets's phrase it. focus, word order, and prosodic phrasing in German double object constructions. In Müller & Sternefeld (2001), (pp. 69–106).

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Massachusetts: MIT Press.

Croft, W. (2001). *Radical Construction Grammar. Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Dekkers, J., van der Leeuw, F., & van de Weijer, J. (Eds.) (2000). *Optimality Theory: Phonology, Syntax and Acquisition*. Oxford: Oxford University Press.

Grimshaw, J. (1997). Projection, heads and optimality. *Linguistic Inquiry*, *28*, 373–422.

Grimshaw, J. & Samek-Lodovici, V. (1998). Optimal subjects and subject universals. In Barbosa et al. (1998), (pp. 193–219).

Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicatures. In D. Schiffrin (Ed.), *Meaning, Form and Use in Context*, (pp. 11–42). Washington: Georgetown University Press.

Jackendoff, R. (1997). *The Architecture of the Language Faculty*. Cambridge, Massachusetts: MIT Press.

Kayne, R. (1994). *The Antisymmetry of Syntax*. Cambridge, Massachusetts: MIT Press.

Kenstowicz, M. & Sohn, H.-S. (1998). Accentual adaptation in north kyungsang korean. Manuscript, MIT and Kyungpook National University. Rutgers Optimality Archive-# 345.

Kuhn, J. (2001). Generation and parsing in optimality theoretic syntax: Issues in the formalization of ot-lfg. In Sells (2001), (pp. 313–366).

Lee, H. (2001a). Markedness and word order freezing. In Sells (2001), (pp. 63–128).

Lee, H. (2001b). *Optimization in Argument Expression and Interpretation: A Unified Approach*. Ph.D. thesis, Stanford University.

Legendre, G. (2001). Masked second-position effects and the linearization of functional features. In Legendre et al. (2001), (pp. 241–277).

Legendre, G., Grimshaw, J., & Vikner, S. (Eds.) (2001). *Optimality Theoretic Syntax*. Cambridge/Mass.: MIT Press.

Müller, G. & Sternefeld, W. (Eds.) (2001). *Competition in Syntax*. Berlin: Mouton de Gruyter.

Pesetsky, D. (1987). Wh-in-situ: Movement and unselective binding. In E. Reuland & A. ter Meulen (Eds.), *The Representation of (In)definiteness*, (pp. 98–129). Cambridge, Massachusetts: MIT Press.

Pesetsky, D. (1997). Optimality theory and syntax: Movement and pronunciation. In Archangeli & Langendoen (1997), (pp. 134–170).

Pesetsky, D. (1998). Some optimality principles of sentence pronunciation. In Barbosa et al. (1998), (pp. 337–384).

Prince, A. & Smolensky, P. (1993/2002). Optimality theory. constraint interaction in generative grammar. Technical report, Rutgers University and University of Colorado at Boulder, Johns Hopkins University (2nd edition only). Manuscript. Republished with minor changes at the Rutgers Optimality Archive, ROA-# 537.

Samek-Lodovici, V. (2002). Prosody-syntax interaction in the expression of focus. Rutgers Optimality Archive, ROA-524.

Schmid, T. & Vogel, R. (submitted). Dialectal variation in German 3-verb-clusters. a surface-oriented ot account. Manuscript, University of Stuttgart and University of Potsdam.

Sells, P. (Ed.) (2001). *Formal and Empirical Issues in Optimality-Theoretic Syntax*. Stanford, California: CSLI Publications.

Tesar, B. & Smolensky, P. (2000). *Learnability in Optimality Theory*. MIT Press.

Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry*, *30*, 219–255.

Vogel, R. (2001). Case conflict in German free relative constructions. An optimality theoretic treatment. In Müller & Sternefeld (2001), (pp. 341–375).

Vogel, R. (2002). Feedback optimisation. Economy and markedness in OT syntax, pt. I. Manuscript, University of Potsdam.

Vogel, R. (to appear). Free relative constructions in OT syntax. In G. Fanselow & C. Féry (Eds.), *Sonderheft Optimality Theory, Linguistische Berichte*. Hamburg: Helmut Buske Verlag.

Wilson, C. (2001). Bidirectional opimization and the theory of anaphora. In Legendre et al. (2001).

Zeevat, H. (2001). The asymmetry of optimality theoretic syntax and semantics. *Journal of Semantics*, *17*, 243–262.