Journal of Semantics 17: 189–216

Some Aspects of Optimality in Natural Language Interpretation

REINHARD BLUTNER
Humboldt University Berlin

Abstract

In a series of papers, Petra Hendriks, Helen de Hoop, and Henriëtte de Swart have applied optimality theory (OT) to semantics. These authors argue that there is a fundamental difference between the form of OT as used in syntax on the one hand and its form as used in semantics on the other hand. Whereas in the first case OT takes the point of view of the speaker, in the second case the point of view of the hearer is taken. The aim of this paper is to argue that the proper treatment of OT in natural language interpretation has to take both perspectives at the same time. A conceptual framework is established that realizes the integration of both perspectives. It will be argued that this framework captures the essence of the Gricean maxims and gives a precise explication of Atlas & Levinson's (1981) idea of balancing between informativeness and efficiency in natural language processing. The ideas are then applied to resolve some puzzles in natural language interpretation.

1 INTRODUCTION

The popularity of Optimality Theory (OT) is notably different in the various fields of linguistics. In phonology it has become the dominant theoretical paradigm. The main reason that OT grew so rapidly in this field is that constraint ranking was silently present in the phonological literature for many years. After the idea was brought from the periphery to the foreground its need in phonology was quite clear.

In syntax, the predominant research tradition has given typically negative answers to the question whether a conflict between constraints is resolved by ranking one constraint over the other. Constraints were assumed to be hard and there is ample evidence that conflicts block the existence of any acceptable output (cf. the discussion in Pesetsky 1997). The recent interest in OT syntax is obvious in the investigation of some non-standard phenomena, especially concerning the interaction between syntax, pronunciation and reference (e.g. Pesetsky 1997). Other motivation came from language typology and from the view that the parser and the grammar are not very different objects. Furthermore, a closer look at the 'absolute' principles has made clear that their violability is actually quite widespread (Speas 1997).

In natural language interpretation the idea of optimization is quite obvious and there is much evidence in favour of competition and constraint ranking in this field. However, the field is rather divergent. Looking at the different conceptions of discourse coherence gives an impression of the heterogeneity of the field. What is essential is a kind of integrative framework that makes it possible to formulate the different conceptions in *one* scientific language and thus to make comparisons between different models transparent. In my opinion, OT is an opportunity for realizing such an integrative framework. However, in its present form OT is insufficient to do this job. So, what we have to do first is to adjust OT to the specific demands of natural language interpretation. Then we can come back to the task of integrating different aspects and different views of natural language interpretation.

In OT it is common to assume three formal components: the **Gen**erator, the **Eval**uator and a system of (ranked) **Con**straints. These components are characterized by three basic assumptions. First, a set of inputs A is assumed. For each input, **Gen** creates a candidate set of potential outputs B. The second assumption is that from the candidate set **Eval** selects the optimal output for that input. The third assumption is that there is a *language particular* ranking of **con**straints from a universal set of constraints. Constraints are absolute and the ranking of the constraints is strict in the sense that outputs that have at least one violation of a higher-ranked constraint can never win over outputs that have arbitrarily many violations of lower-ranked constraints (cf. Prince & Smolensky 1993; Kager 1999).

Each of these three assumptions has to be adjusted or revised in order to satisfy the demands of natural language interpretation. With respect to **Gen**, I think, it is best to take a dynamic picture of natural language semantics and to describe it in terms of a context change semantics. This adjustment is especially important in order to deal with the context dependency of natural language interpretation (e.g. Kamp & Reyle 1993). Next, consider **Eval**. The direction of optimization is usually taken unidirectional (from A to B, where the elements of A sometimes are called inputs and the elements of B outputs). One of my main arguments is that in the case of interpretation it is inevitable to have bidirection of optimization (from A to B and from B to A). Both directions are not independent of each other; instead, they should be interrelated in a particular way. Third, with regard to **Con** we have to acknowledge the role of graded constraints. Graded constraints also appear in other domains, for example in phonology (cf. Prince & Smolensky 1993; Boersma 1998). However, in natural language interpretation the role of graded constraints seems to be much more important than in other domains. Another point is that in natural language interpretation the relevant pragmatic constraints are always ranked universally within the set of pragmatic constraints. As a consequence, typological differences between languages are not triggered by a reranking of the constraints **within** the pragmatic domain. Instead, typological effects are triggered—among other things—by variations that concern the relative importance of pragmatic constraints with regard to other types of constraints. Choi (1996) supports this point in an indirect way by comparing scrambling phenomena in German and Korean.

The paper is structured as follows. In section 2 some arguments are put forward as to why bidirection of optimization is of central importance when we try to apply OT to natural language interpretation. Section 3 introduces my proposals for a proper treatment of optimality in natural language interpretation. The starting point is the context change potential of an (underspecified) expression which is described as a relation between input and output contexts. The effect of optimality is simply to constrain this relationship in a way which both involves optimization for interpretation and optimization for production. In section 4 the general framework is put in concrete terms by modelling contexts as DRSs. It is demonstrated that van der Sandt's/Geurts' projection mechanism for presuppositions can be reconstructed and extended as a consequence of the present form of OT.

2 TWO PERSPECTIVES OF OPTIMALITY

De Hoop & de Swart (1998), Hendriks & de Hoop (to appear), and de Hoop (2000) applied OT to sentence interpretation. These authors argue that there is a fundamental difference between the form of OT as used in syntax on the one hand and its form as used in semantics on the other. Whereas in the former case OT takes the point of view of the speaker (production perspective), in the latter case the point of view of the hearer is taken (comprehension perspective).[1]

This idea is an important one and I think most of the existing analyses conform to it. Moreover, the picture can be extended to OT phonology and morphology. For example, in phonology **Gen** clearly takes the production perspective and creates a candidate set of potential outputs (=speech sounds as they occur in utterances) for a given input (=speech sounds as they occur in the mental lexicon). From the candidate set, **Eval** selects the best (optimal) output for that input. A similar picture can be found in OT morphology (e.g. Bresnan, to appear). Here the input

[1] By using the terms 'comprehension' and 'production' we do not refer to performance but rather to abstract functions in a mathematical sense that pair certain pairs of representations (cf. Smolensky 1996).

represents language-independent 'content' in the multidimensional space of possible grammatical and lexical contrasts and **Gen** enumerates a set of concrete realizations of the input that are available across languages (expressing the 'content' with varying fidelity).

However, the one way tableau typically assumed in phonology may be insufficient. One reason for this shortage has to do with the nature of the input under OT. In contrast to standard generative phonology, where numerous constraints were imposed on the input, in OT constraints on the input are typically lacking. In principle, the set of inputs to the grammars of all languages is assumed to be the same (*richness of the base*). As a consequence, in many cases it is easy to construct multiple inputs that converge on a single output. Which of the multiple inputs should be selected? This question is important when we assume that the relevant inputs must be stored somewhere in the mental lexicon. The economy of the lexicon requires that corresponding inputs must be selected careful. Prince & Smolensky (1993: section 9) introduced an algorithm called *lexicon optimization* (further developed by Itô, Mester, & Padgett 1995) which optimizes the inputs. The algorithm examines the constraint violations incurred by the winning output candidate corresponding to each competing input. The input–output pair with the fewest violations is selected as the optimal pair. Thus, lexicon optimization works from the inputs A to the outputs B and back from B to A. As a consequence, the 'input' set A is restricted in an indirect way, by means of the system of ranked constraints and the possible outputs.

OT syntax is another case where the production perspective is taken exclusively. It optimizes syntactic structure with respect to a semantic input. Now we have to notice human sentence parsing as a related area in which optimality has always been assumed. According to the nature of parsing, in this case the comprehension perspective comes in. Consequently, the parser optimizes underlying structures with respect to a surface input. Gibson & Broihier (1998) and Fanselow, Schlesewsky, Cavar, & Kliegl (1999) have shown that parsing preferences can be explained in this way. Furthermore, Fanselow, Schlesewsky, Cavar, & Kliegl (1999) have tried to demonstrate that the same constraints seem to be used both in OT syntax and parsing. If this it right, it demonstrates that both directions of optimization are relevant. OT syntax normally ignores the phenomenon of syntactic ambiguities and does not try to explain the preferences for the different readings that suggest itself (cases in point are quantifier scope and PP-attachment). I see it as an opportunity for OT syntax to explain the relevant preferences with the help of syntactic constraints, which are motivated independently. If we consider optimality under the production perspective exclusively, we lose this opportunity to give a *syntactic* explanation for the preferences. This does not exclude the relevance of pragmatic factors that arguably interact with the syntactic factors.

Now let us address natural language interpretation. Ambiguity, polysemy, and other forms of flexibility are much more obvious and manifested in a much broader way in this area than in the realm of syntax. The assumption that OT in sentence interpretation takes the point of view of the hearer is mainly motivated by this observation and the aim to explain the interpretive preferences. Using this perspective a mechanism for preferred interpretations is constituted that provides insights into different phenomena of interpretations, such as the determination of quantificational structure (Hendriks & de Hoop, to appear), nominal and temporal anaphorization (de Hoop & de Swart 1998), and the interpretational effects of scrambling (de Hoop 2000). However, I think there are reasons demonstratnig this design of OT to be inappropriate and too weak in a number of cases. The reasons have to do with the fact that **Gen** can pair *different* forms with one and the same interpretation. The existence of such alternative forms may raise blocking effects that strongly affect what is selected as the preferred interpretation. It is not difficult to see that the arguments for a bidirectional view in syntax and the arguments for a bidirectional view in interpretation are complementary. In the case of syntax, we cannot explain *interpretative preferences* when we take the production perspective alone. In the case of semantics/pragmatics we cannot explain *blocking effects* when we take the comprehension perspective alone.

Blocking effects are essential for the explanation of pragmatic anomalies. This may be illustrated with an example. Consider the well-known phenomenon of 'conceptual grinding', whereby ordinary count nouns acquire a mass noun reading denoting the stuff the individual objects are made of, as in *Fish is on the table* or *Dog is all over the street.* One of the essential factors that restrict the grinding mechanism is lexical blocking. For example, in English the specialized mass terms *pork, beef, wood* usually block the grinding mechanism in connection with the count nouns *pig, cow, tree.* This explains the contrasts given in (1).

(1) a. I ate pork/?pig
    b. Some persons are forbidden to eat beef/?cow
    c. The table is made of wood/?tree

Blocking effects need not be absolute. Instead, they may be cancelled under special contextual conditions. Nunberg & Zaenen (1992) give the following example of what they call *deblocking*:

(2) Hindus are forbidden to eat cow/?beef

They argue that what makes *beef* odd here is that the interdiction concerns the status of the animal as a whole, and not simply its meat. That is, Hindus are forbidden to eat beef only because it is cow-stuff. Copestake & Briscoe (1995) provide further examples that substantiate this claim.

The simplest explanation for blocking (and also deblocking) is a bidirectional OT that takes into account the production perspective. An expression is blocked with regard to a certain interpretation if this interpretation can be generated more economically by an alternative expression. Linguistic and contextual factors can trigger deblocking in case they reverse the corresponding cost values (cf. Copestake & Briscoe 1995; Blutner 1998).

The binding behaviour of pronominal expressions gives another illustration for the importance of blocking in natural language interpretation.

(3) a. John_i washes himself_i
b. *John_i washes him_i
c. John_i expected Mary to wash him_i

In (3b) the coreferential reading is impossible because this interpretation is blocked by the form (3a) which is assumed to be more cheaply generated (because of a weak constraint saying 'bound NPs are marked reflexive'). In (3c) this blocking effect is cancelled out by a higher-ranked constraint 'A reflexive must be bound locally' (Burzio 1998). The version of (3c) with a reflexive will now be taken to violate this constraint, while the one with the pronoun only violates the lower-ranked constraint 'bound NPs are marked reflexive', thus representing the optimal candidate.

Appreciating the basic findings of Petra Hendriks, Helen de Hoop and Henriëtte de Swart concerning the selection among interpretations, the conclusion can only be that we have to consider bidirectional optimization. This appears to be almost a conceptual necessity.

A careful argument in favour of bidirectionality has to take into account the important distinction between a semantic representation (=formal meaning) and an interpretation (content). If we identify semantic representations and interpretational content, then we simply have to state that a bidirectional OT is established by combining OT syntax and OT semantics

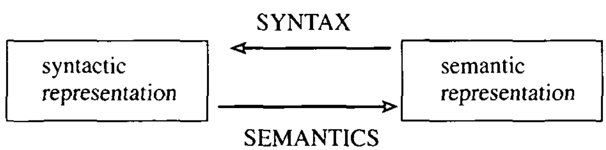


Figure 1 Syntax and semantics as the two directions of bidirectional OT

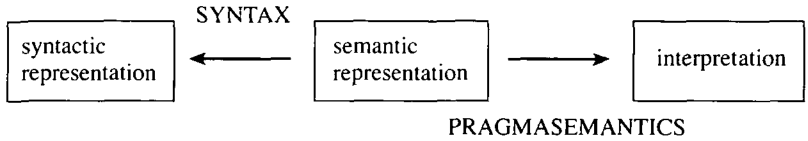


Figure 2 The two directions of optimization in a model without bidirection

(see Figure 1). OT semantics takes syntactic representations as inputs and results in optimal semantic outputs, and OT syntax takes semantic representations as inputs and results in optimal syntactic outputs. To say that we need bidirectionality is then simply to say that we need OT syntax **and** OT semantics. Presumably, this is the view taken by the pioneers of OT semantics.

There are different schools of linguistics which consider the distinction between formal meaning and interpretational content as a very important issue. For example, Bierwisch (1983, 1996) proposed his *two-level semantics*, Carston (1998) made a similar point from the perspective of *relevance theory*, and many people in computational linguistics have a related distinction based on the idea of underspecification (e.g. van Deemter & Peters 1996). Assuming this distinction could lead us to an architecture combining the ideas of optimal production and optimal interpretation in a way that does not make use of bidirection (Figure 2). It is not difficult to see that this architecture is unable to explain the blocking of interpretations in the general case. It only describes the blocking of interpretations just for those cases where the corresponding semantic representations are blocked. The example of 'conceptual grinding' and other phenomena within the realm of lexical pragmatics (cf. Blutner 1998) suggest that one and the same semantic representation may be connected with a variety of different interpretations. Nevertheless, certain interpretations can be blocked without blocking the corresponding semantic representations.

It is not difficult to suggest an architecture that doesn't suffer from these shortcomings. It is shown in Figure 3. Here we have to consider two modes of bidirection—one for relating syntactic and semantic representations and one for relating semantic representations and interpretations. It goes without saying that this architecture does not really **conflict** with the

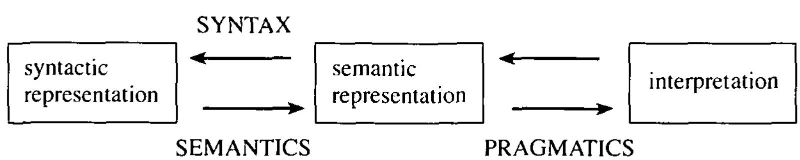


Figure 3 A model with two modes of bidirection

ideas of the pioneers of OT semantics. Instead it broadens their view in a straightforward way.

Not surprisingly, it is rather unclear sometimes which phenomenon should be treated within which mode of bidirection. Consider the case of binding phenomena. Building on Burzio (1989), Colin Wilson (1998) develops a theory of anaphora incorporating two types of competition. Assuming the interface between syntax and semantics to have a particular 'direction', Wilson takes both directions into account—the direction that maps from semantic structures to syntactic ones and the opposite direction that maps from syntactic structures to semantic ones. Clearly, Wilson's account refers to the mode of bidirection shown on the left-hand side of Figure 3. In contrast, there is Levinson's pragmatic theory of anaphora (e.g. Levinson 1987), which can be seen as operating in the pragmatic mode of bidirection (right-hand side of Figure 3). It is not the aim of this paper to judge which decision is the better one.

Independent of the position we take with regard to the distinction between meaning and interpretation, the advantage of the bidirectional view becomes clear now: it integrates interpretational preferences and blocking effects and it keeps OT simple: 'What is best expressed as a generation principle is expressed as a generation principle, what is best expressed as an interpretation principle is expressed as an interpretation principle' (Zeevat, this volume).

Under the present perspective of integrating production and comprehension optimality we can account both for ineffability and for pragmatic anomaly. The first case occurs when the optimal production can be triggered more efficiently by an alternative interpretational input. The second case occurs when the optimal interpretation can be expressed more efficiently by an alternative form.[2]

The final remark has to do with the foundation of OT in Harmony Theory. Harmony Theory is a formalism which abstracts away from the details of connectionist networks and seeks to find out general mathematical techniques for analysing classes of connectionist networks (Prince & Smolensky 1993; Smolensky 1986). One essence of Harmony Theory is its founding on a two-layer scheme which allows a combination of simplicity with uniformity. On the lower layer we find *representational nodes* that encode the different kinds of information involved in language processing

[2] Another nice example where a bidirectional competition technique can help to explain empirical generalizations is discussed by Lee (2000). Based on the constraints assumed by Choi (1996), Lee shows that a bidirectional model can explain some types of 'freezing effects' concerning the word order in German and Korean (looking at sentences with ambiguous case marking). For further examples and references, see Kuhn (2000) and the web page of Bresnan: http://www-lfg.stanford.edu/lfg/bresnan/.

(phonological, morphological, syntactic, semantic). On the upper level we find *knowledge nodes* that are hidden units that encode certain 'patterns' that relate particular configurations of representational units. A connectionist network is a dynamical system that is controlled by a certain Ljapunov function. When activation dynamically spreads off, this function always decreases or remains constant. In other words, harmony theory says that starting from any incomplete representational vector, this vector is always completed in a minimalistic/optimal way.

Harmony theory does not say that the different optimizations converge when we start with different parts of a lucid representational vector. The theory says only that one and the same Ljapunow function (=system of ranked constraints in OT) can be used when the system operates like a hearer (starting with a natural language form and ending with an interpretation) or when it operates like a speaker (starting with an activated interpretation and ending with a form). The theory does not say that we come back to the original expression when we execute both operations in succession.

Everyone can describe numerous situations in which he was unable to produce what he understands. More drastically, the phenomenon of aphasia illustrates possible asymmetries in production and comprehension (e.g. Jakobson 1941/1968). A related asymmetry is found in language acquisition. It is well known that children's abilities in production lag dramatically behind their abilities in comprehension. In overcoming this lag, a kind of bootstrap mechanism seems to apply that depends crucially on the *robustness* of comprehension, possibly by using a technique called *robust interpretative parsing* (Smolensky 1996; Tesar & Smolensky 2000). Consequently, when it comes to relate the two perspectives within a bidirectional OT, we have to acknowledge the close interrelation between them in the OT learning algorithm. In summary, harmony theory *per se* does not give any argument in favour of bidirection. Instead, the arguments are coming from OT learning theory. I will come back to this important conceptual point in the next section.

3 AN INTEGRATIVE FRAMEWORK

In this section an attempt is made to integrate optimal interpretation and optimal production. A look at the area of pragmatics seems to be useful since an analogous optimality metric plays an indispensable role there. The Gricean conversational maxims are widely recognized as a (rather informal) expression of this metric. With Zipf (1949) as a forerunner we have to acknowledge two basic and competing forces, one force of unification, or Speaker's economy, and the antithetical force of diversification, or Auditor's economy. The two opposing economies are in extreme conflict, and we have to look for an optimal way to resolve this conflict.

An important step in reformulating and explicating the Gricean framework has been made by Atlas & Levinson (1981) and Horn (1984), who have tried to clarify the consequences of these opposing economies. Taking Quantity as a starting point, they distinguish between two principles, the Q-principle and the I-principle (termed R-principle by Horn 1984). The I-principle can be seen as the force of unification minimizing the Speaker's effort, and the Q/R-principle can be seen as the force of diversification minimizing the Auditor's effort. Simple but informal formulations of these principles are as follows:

(4) *Q-principle*: Say as much as you can (given I) (Horn 1984: 13).
Do not provide a statement that is informationally weaker than your knowledge of the world allows, unless providing a stronger statement would contravene the I-principle (Levinson 1987: 401).
I-principle: Say no more than you must (given Q) (Horn 1984: 13).
Say as little as necessary, i.e. produce the minimal linguistic information sufficient to achieve your communicational ends (bearing the Q-principle in mind) (Levinson 1987: 402).
Read as much into an utterance as is consistent with what you know about the world (Levinson 1983: 146–7).

Obviously, the Q-principle corresponds to the first part of Grice's quantity maxim (*make your contribution as informative as required*), while it can be argued that the countervailing I-principle collects the second part of the quantity maxim (*do not make your contribution more informative than is required*), the maxim of relation and possibly all the manner maxims.

In a slightly different formulation, the I-principle seeks to select the most coherent interpretation, and the Q-principle acts as a blocking mechanism and blocks all the outputs that can be derived more economically from an alternative linguistic input (for a detailed discussion see Blutner 1998). This formulation makes it quite clear that the Gricean framework can be understood in a bidirectional optimality framework which integrates production and comprehension optimality. At first glance, using a bidirectional competition technique can be seen merely as establishing the very same ideas as presented in Blutner (1998) using a more widely acknowledged and well-known basis. However, that is not the whole story. We have to acknowledge that the framework of OT gives us a much wider perspective on relating natural language comprehension, language acquisition (Tesar & Smolensky 2000) and language change (e.g. Haspelmath 1999). Furthermore, there are interesting mathematical results concerning the computational capacity of OT systems (see Kuhn 2000 for further references). Taking the broader perspective and the more rigorous formalization, the use of OT may give the enterprise of Radical Pragmatics in general and Lexical Pragmatics in particular a new impulse.

With the Gricean maxims as **Eval**, we have to make more explicit now the status of **Gen**. Following current trends in semantics, we see the formal meaning of a natural language expression A as its context change potential (e.g. Heim 1982; Kamp 1981; Kamp & Reyle 1993; Groenendijk & Stokhof 1991). It describes the way A (or better, the semantic form sem(A) that is associated with A) updates the current context σ leading to a new context τ. In standard dynamic semantics the context change potential is assumed to be a function, with the argument of the function usually written to the left: $\sigma[\text{sem}(A)] = \tau$. Taking into account that the semantics is highly underspecified (e.g. Reyle 1993) and that it seldom specifies a definite outcome, we assume that the context change potential is a relational notion. If τ is one of the potential outcomes of updating σ with sem(A), this is written as $\sigma[\text{sem}(A)]\tau$. The Generator $\mathbf{Gen}_\sigma$ is now identified with the set of input-output (form-interpretation) pairs $\langle \text{sem}(A), \tau \rangle$ such that τ is a potential result of updating σ with sem(A); more formally:

(5) $\mathbf{Gen}_\sigma = \{\langle \text{sem}(A), \tau \rangle \colon \sigma[\text{sem}(A)]\tau\}$

For convenience, we will simply write A instead of sem(A) from now on.

The effect of the Gricean maxims is simply to constrain this relation in a particular way, and we have already given some initial motivation that this constraint can be formulated best in a bidirectional OT framework. In OT there is a cost function (harmony function) that evaluates the elements of the generator. For the present aims it is sufficient to assume an ordering relation $\succ$ (*being more harmonic, being more economical*) that ranks the elements of the Generator.[3]

Now the following formulation of the Q and the I-principle comes immediately to mind and brings us to a bidirectional optimality view:

(6) *Bidirectional OT (strong version)*

(Q) $\langle A, \tau \rangle$ satisfies the Q-principle iff $\langle A, \tau \rangle \in \mathbf{Gen}_\sigma$ and there is no other pair $\langle A', \tau \rangle$ such that $\langle A', \tau \rangle \succ \langle A, \tau \rangle$

[3] Being more pedantic, we should write $\succ_\sigma$ in order to indicate the dependence on the actual context σ. We can drop the index because here and in the following we assume the actual context to be fixed.

(I) $\langle A, \tau \rangle$ satisfies the I-principle iff $\langle A, \tau \rangle \in \mathbf{Gen}_\sigma$ and there is no other pair $\langle A, \tau' \rangle$ such that $\langle A, \tau' \rangle \succ \langle A, \tau \rangle$

$\langle A, \tau \rangle$ is called *optimal* iff it satisfies both the Q-principle and the I-principle.[4]

Obviously, a pair $\langle A, \tau \rangle$ satisfies the Q principle just in case A is an optimal production that can be generated starting with τ. On the other hand, a pair $\langle A, \tau \rangle$ satisfies the I-principle just in case τ is an optimal outcome of interpreting A. Seeing both principles as being part of the real mechanism of natural language comprehension, the I-principle can be considered as a sub-mechanism for finding out preferred interpretations, and the Q-principle can be considered as an (absolute) blocking mechanism that suppresses the interpretations that are connected more economically with an alternative form.

In standard OT the ordering relation between elements of the generator is established via a system of ranked constraints. These constraints are typically assumed to be output constraints, i.e. they may be either satisfied or violated by an output form. In the bidirectional framework just presented, changing the perspectives is possible. This means that an output under one perspective can be seen as an input under the other perspective. Therefore, it is plausible to assume output *and* input constraints. However, we should avoid (relational) constraints that refer to inputs and outputs simultaneously. Seeing the input as a linguistic form that conveys phonological, syntactic, and semantic information, input constraints are typically markedness conditions that evaluate the 'harmony' of the form. On the other hand, the output (i.e. the resulting context τ) is evaluated by constraints that determine its coherence and informativeness (with regard to the initial context σ).

Let me now give a very schematic example in order to illustrate some characteristics of the bidirectional OT (labelled *strong version* in order to discriminate it from a *weak version* introduced later). Assume that we have two constraints called F and C. F is a constraints on linguistic forms and collects the effects of linguistic markedness. C is a constraint on resulting contexts and refers to coherence and informativeness. There is no reason to introduce a ranking between F and C. Let us assume two forms A_1 and A_2 which are semantically equivalent. That means $\mathbf{Gen}_\sigma$ associates the same relations of context change with them. With σ as initial context, let us assume the possible outcomes are τ_1 and τ_2. Further, we assume that no other form updates σ to one of these outcomes. Let us stipulate that A_1 satisfies F but not A_2 and that τ_1 satisfies C but not τ_2. That makes the form

[4] In terms of game theory, the solution concept that underlies the formulation of (strong) optimality is that of a 'Nash Equilibrium' (see Dekker & van Rooy, this volume).

A_2 less well-formed than the form A_1 and the resulting context τ_2 more complex than the resulting context τ_1. The bidirectional view can be demonstrated by the following tableau, where two *super-columns* are introduced, one for each result of context change.

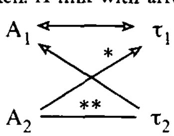
(7)

Forms		F	C		F	C
A_1	☞ ➳			☞		*
A_2	➳	*			*	*
Interpretations		τ_1			τ_1	

I use Smolensky's (1996) repertoire of symbols here: ☞ indicates the optimal candidate when the production perspective is taken (find an optimal expression starting with τ_i) and ➳ indicates the optimal candidate when the comprehension perspective is taken (find an optimal interpretation starting with A_i). *Super-optimal* pairs are those that are production and comprehension optimal. This is indicated by the simultaneous occurrence of ☞ and ➳. The tableau shows that only the form A_1 survives, with τ_1 as its only interpretative outcome. Obviously, the form A_2 is blocked in all its (semantically admissible) interpretations.[5]

The scenario just installed describes the case of *total blocking* where some forms (e.g. *furiosity, *fallacity) do not exist because others do (fury, fallacy). However, blocking is not always total but may be partial. According to Kiparsky (1982), *partial blocking* is realized in the case where the special (less productive) affix occurs in some restricted meaning and the general (more productive) affix picks up the remaining meaning (consider examples like *refrigerant – refrigerator, informant – informer, contestant – contester*). To handle these and other cases Kiparsky (1982) formulates a general condition *Avoid Synonymy*. Working independently of the Aronoff–Kiparsky line, McCawley (1978) collects a number of further examples demonstrating the phenomenon of partial blocking outside the domain of derivational and inflectional processes. For example, he observes that the distribution of

[5] Zeevat (personal communication) has proposed using pictures of the following kind, where arrows indicate the optimal candidate that arises when the indicated direction of optimization is taken. A link with arrows in both directions indicates a *super-optimal* pair.



productive causatives (in English, Japanese, German, and other languages) is restricted by the existence of a corresponding lexical causative. Whereas lexical causatives (e.g. (8a)) tend to be restricted in their distribution to the stereotypical causative situation (direct, unmediated causation through physical action), productive (periphrastic) causatives tend to pick up more marked situations of mediated, indirect causation. For example, (8b) could have been used appropriately when Black Bart caused the sheriff's gun to backfire by stuffing it with cotton.

(8) a. Black Bart killed the sheriff
 b. Black Bart caused the sheriff to die

Typical cases of total and partial blocking are not only found in morphology, but in syntax and semantics as well (cf. Atlas & Levinson 1981; Horn 1984; Williams 1997). The general tendency of partial blocking seems to be that 'unmarked forms tend to be used for unmarked situations and marked forms for marked situations' (Horn 1984: 26)—a tendency that Horn (1984: 22) calls '*the division of pragmatic labour*'.

There are two principal possibilities avoiding total blocking within the bidirectional OT framework. The first possibility is to make some stipulations concerning **Gen** excluding equivalent semantical forms. Such a case is demonstrated in (9):

(9)

Forms		F	C		F	C
A_1	☞ ➵			[redacted]		
A_2	➵	*		☞	*	*
Interpretations		τ_1			τ_1	

In this case the unmarked form A_1 is stipulated to be used for the unmarked situation only. (This seems plausible when we assume the child learns the meaning of *kill* in stereotypical, unmarked situations). The interpretation of the marked form A_2 remains open. Unfortunately, the bidirectional OT described in (6) does not select any situation for A_2. Starting with τ_2, expressive optimization selects A_2, as desired. However, we do not come back to the marked situation τ_2 when the inverse perspective (interpretative optimization) is taken. Instead, the unmarked situation τ_1 is selected. Consequently, there is no output that is paired super-optimal with A_2. That means, A_2 is blocked in all interpretations.

The only possibility to account for Horn's division of pragmatic labour is to stipulate it as a property of the Generator. This is indicated by the following tableau:

(10)

Forms		F	C		F	C
A_1	☞ ⇒					
A_2				☞ ⇒	*	*
Interpretations		τ_1			τ_1	

Obviously, this solution is completely ad hoc, and we should look out for an alternative solution.[6]

The bidirectional OT we have considered until now is a very strong and absolute one. We have assumed (i) that an input-output pair $\langle A, \tau \rangle$ is *super-optimal* just in case τ is optimal for A and A is optimal for τ, and (ii) that the bidirections of optimization are independent of each other. This means that the results of optimization under one perspective are not assumed to influence which structures compete under the other perspective.

Our initial motivation for developing a bidirectional OT was the formulation of the Gricean maxims in Radical Pragmatics (Atlas & Levinson 1981; Horn 1984). Already the informal formulations given in (4) make it completely clear that we need a formalization where bidirections of optimization refer to each other. Such a formalization has been given in Blutner (1998):

(11) *bidirectional OT (weak version)*

(Q) $\langle A, \tau \rangle$ satisfies the Q-principle iff $\langle A, \tau \rangle \in \mathbf{Gen}_\sigma$ and there is no other pair $\langle A', \tau \rangle$ *satisfying the I-principle* such that $\langle A', \tau \rangle \succ \langle A, \tau \rangle$

(I) $\langle A, \tau \rangle$ satisfies the I-principle iff $\langle A, \tau \rangle \in \mathbf{Gen}_\sigma$ and there is no other pair $\langle A, \tau' \rangle$ *satisfying the Q-principle* such that $\langle A, \tau' \rangle \succ \langle A, \tau \rangle$

$\langle A, \tau \rangle$ is called *super-optimal* iff it satisfies both the Q-principle and the I- principle.[7]

[6] As suggested by an anonymous referee, there is a further argument that shows that it is problematic to have hard constraints for excluding total blocking. In fact, a sentence like (8a) CAN be used in situations where Black Bart caused the Sheriff's gun to backfire it with cotton. This possibility is excluded when hard constraints are used as in (9) and (10).

[7] Recently, Gerhard Jäger (Jäger 1999; see also Jäger & Blutner to appear) has presented a more transparent formulation of bidirectional OT:

I call this variant of the bidirectional OT the *weak* version. The important point is that the structures that compete in one perspective of optimization are constrained by the outcomes of the other perspective and vice versa. The purpose of this kind of recursive dependence can be demonstrated by coming back to our original example which leads now to the following tableau:

(12)

Forms		F	C		F	C
A_1	☞ ➳					*
A_2		*		☞ ➳	*	*
Interpretations		τ_1			τ_1	

Let us take first the comprehension perspective starting with A_1. The structures that compete are $\{\tau_1, \tau_2\}$ (the marked form A_2 does not block any of them). From the fact that τ_1 is less expensive (more stereotypical) than τ_2 it follows that the little arc ➳ has to select τ_1. Now take the production perspective starting with τ_1. An analogous argument shows that the little hand ☞ selects A_1. Consequently, the pair $\langle A_1, \tau_1 \rangle$ is *super-optimal*—just as in tableau (7) where we discussed the *strong view*. Next consider the comprehension perspective starting with A_2. In this case the structures that compete are restricted to the singleton $\{\tau_2\}$ since the unmarked form A_1 blocks τ_1, and we get that the little arc ➳ has to select τ_2. An analogous argument applies to the production perspective starting with τ_2. In this case the competition set is restricted to the singleton $\{A_2\}$, and the little hand ☞ selects A_2. In contrast to the *strong* view, now the pair $\langle A_2, \tau_2 \rangle$ comes out as super-optimal as well. And this demonstrates that the *weak* view can

$\langle A, \tau \rangle$ is super-optimal iff $\langle A, \tau \rangle \in \mathbf{Gen}_\sigma$ and
(Q) there is no super-optimal $\langle A', \tau \rangle < \langle A, \tau \rangle$
(I) there is no super-optimal $\langle A, \tau' \rangle < \langle A, \tau \rangle$.

Jäger has shown that there is a unique super-optimality relation in case $<$ is well founded. Furthermore, this formulation of super-optimality is equivalent to that presented in (11) if $<$ satisfies transitivity. Jäger's results demonstrate that the circularity inherent in definition (11) is an apparent one only. Suppose the preference relation $<$ as well founded, then both the definition (11) and Jäger's definition of super-optimality come out as sound recursive definitions (cf. also Dekker & van Rooy, this volume).

Does the recursive variant of bidirectionality (i.e. weak bidirection) extend the computational capacity of the generator and, if yes, in which way? These are important but largely unsolved problems even for unidirectional OT. (For some interesting results concerning the system OT-LFG, cf. Kuhn 2000). Gerhard Jäger (p.c.) has a proof that under the same conditions that are assumed in Karttunen (1998), weak bidirection does not extend the generative capacity of the generator.

account for the good old idea that unmarked forms tend to be used for unmarked situations and marked forms for marked situations.

One consequence of the *strong mode* of optimization in (6) can be summarized as follows: *What we produce we are able to understand adequately and what we understand we are able to produce adequately.* At least the second part of this consequence is clearly false when we consider children's ability in natural language production, which lags dramatically behind their ability in comprehension. Smolensky (1996) has demonstrated that OT gives an plausible explanation for this lag. OT predicts that in comprehension relatively marked forms can be understood appropriately. However, when we consider generation, then highly unmarked forms are produced that significantly differ from the initial forms. The lag between comprehension and production is overcome by learning. According to the OT learning theory (Smolensky 1996; Tesar & Smolensky 2000), learning results in a state of the system that satisfies the demands of strong bidirection.

It is easy to prove that a pair that is *optimal* (strong bidirection, cf. (6)), is *super-optimal* (weak bidirection, cf. (9)) as well. However, weak bidirection gives a chance to find additional *super-optimal* solutions. This is demonstrated by tableau (12). Is it possible to give a natural interpretation for these additional solutions? I want to propose the idea that these additional solutions are due to the flexibility and ability to learn which the weak formulation alluded to.

In my opinion, the weak version of the bidirectional OT can be taken to describe the possible outcomes of self-organization before the learning mechanism has fully realized the equilibrium between productive and interpretative optimization. Jäger (1999) and Dekker & van Rooy have proposed algorithms that update the ordering (preference) relation $\succ$ such that (i) optimal pairs are preserved and (ii) a new optimal pair is produced if and only if the same pair was super-optimal at earlier stages. Consequently, we can take the solutions of weak bidirection to be identical with the solutions of strong bidirection considering all the systems that result from updating the ordering relation. Arguably, updating the ordering relation in the style of Jäger describes a kind of self-organization which is very close to certain mechanisms of self-organization in language change. This point may be clarified when we (re)consider the principle of iconicity (called 'the division of pragmatic labour' within the domain of pragmasemantics). This principle can be proven to result from weak bidirection (ask Gerhard for the proof). In the school of *natural morphology* (for references cf. Wurzel 1998), the same principle plays an important role in describing the direction of language change.

Constructional iconicity: A semantically more complex, derived morphological form is unmarked regarding constructional iconicity if it is symbolized formally more costly than its semantically less complex base; it is the more marked, the stronger its symbolization deviates from this (Wurzel 1998: 68).

Analogies of this kind give substance to the claim that *weak bidirection* can be considered as a principle describing (in part) the direction of language change: super-optimal pairs are tentatively realized in language change. This relates to the view of Horn (1984) who considers the Q principle and the I principle as diametrically opposed forces in inference strategies of language change.

4 PRESUPPOSITION PROJECTION

In the previous section we have outlined two general ideas that determine the shape of **Gen** in natural language interpretation: underspecification and dynamic semantics. Within the realm of underspecification we can discriminate between structural underspecification and lexical underspecification. Structural underspecification is related, for example, to scope, ellipsis, and presupposition. Lexical underspecification, on the other hand, relates to polysemy, metonymy, and other aspects of the 'Generative Lexicon'. Although it is seldom made completely explicit in OT, the choice of a particular representational format is unavoidable in order to be give a sound formulation of the constraints and their ranking. With regard to the representational format, we will proceed by modelling contexts as DRSs. Moreover, the initial DRSs of presupposition-inducing expressions are treated in the particular framework of van der Sandt (1992) and Geurts (1995). This framework combines the idea of dynamics with the aspect of underspecification that relates to presupposition projection.

The aim of this section is to demonstrate that van der Sandt's/Geurts' projection mechanism for presuppositions can be reconstructed (in important aspects) and improved (in secondary aspects) as a consequence of the I-principle. Moreover, it can be explained why accommodation is sometimes blocked. This is an important consequence of the Q-principle, and its integration realizes an effective extension of the van der Sandt/ Geurts proposal.

As usual, we consider a DRS K as a pair $\langle U(K), Con(K) \rangle$, where $U(K)$ is a set of reference markers and $Con(K)$ is a set of DRS-conditions. If P is an n-place predicate, and $x_1, \ldots, x_n$ are reference markers, then $P(x_1, \ldots, x_n)$ is a *simple DRS-condition*. If K and K′ are DRSs, then $\neg K$, $K \vee K'$, $K \Rightarrow K'$ are *(complex) DRS-condition* (cf. Kamp & Reyle 1993; Kadmon 1990; Geurts 1995, 1999).

In order to account for *presupposition inducers* we introduce a further type of complex DRS-conditions: conditions of the form B/K, where K is a DRS and B is a DRS-condition. Conditions B/K have a special status and are called slash-conditions. They induce presuppositions and mark them as material *behind the slash*. Though not identical, this notation is very similar to that of Geurts (1995). The role of slash conditions is to indicate that a presupposition may be bound or accommodated in any DRS that subordinates the DRS in which it originates. Since the structural position where the presupposition is resolved/accommodated is not specified semantically, an element of structural underspecification is introduced into the whole framework. More formally, let σ and τ be ordinary DRSs and sem(A) be a DRS that may contain slash conditions (introducing presupposed material). Then the idea can be expressed by the following notion of context change:

(13) σ [sem(A)] τ just in case τ is the result of merging[8] σ with the result of projecting the presupposed material of sem(A) such that the resulting DRS is a proper one (it may not contain any free reference markers).[9]

Using the conception of **Gen** as defined in (5), the formulation in (14) results where the Generator is considered for a specific input form A:

(14) $\mathbf{Gen}_\sigma$ (A) = {τ: τ is the result of merging σ with the result of projecting the presupposed material of sem(A) such that the resulting DRS is a proper one}

The part of the projected DRS that factors with part of the superordinated DRS/initial context (σ) will be called *bound* (or *resolved) material*; the part that does not factor will be called *accommodated material*. For convenience, in the corresponding DRSs, the part of the presupposition which counts as *bound* when projected is *underlined*, and the part which has to be *accommodated* is *underlined twice*.

[8] *DRS-merge* (cf. Geurts 1995): If $\boldsymbol{K}$ is a set of DRSs, then $\oplus \boldsymbol{K} = \langle \cup_{K \in \boldsymbol{K}} \mathrm{U}(K), \cup_{K \in \boldsymbol{K}} \mathrm{Con}(K) \rangle$.

[9] A necessary condition is that presupposed material projects to a *DRS* that subordinates the origin position.

Subordination (cf. Geurts 1995): $\leq$ is the smallest preorder (transitive, reflexive) for which all of the following hold, for any K, K′, K″:

a. If $\neg K' \in$ Con(K), then $K \leq K'$
b. If $K' \vee \mathbf{K}'' \in$ Con(K), then $K \leq K'$ and $K \leq K''$
c. If $K' \Rightarrow K'' \in$ Con(K), then $K \leq K' \leq K''$
d. If $B/K' \in$ Con(K), then $K \leq K'$
(Read $K' \leq K$ as K' *subordinates* K).

Let us give two simple examples. In (15) a conditional A is given and its semantic form sem(A) is indicated. With regard to an initial context that is empty (Ø) three projections of the presupposed material are possible. They are indicated by τ_1, τ_2, τ_3 and refer to what is usually called local, intermediate, and global accommodation, respectively. Binding is not possible in these situations.

(15) A: If Peter has a dog, then his cat is gray

$$\text{sem}(A) \;=\; [\;:[\text{x: dog(x), have(Peter, x)}] \Rightarrow [\;:\text{gray(y)} \;/\; [\text{y: have(Peter, y), cat(y)}]]]$$

$$\text{Gen}(A) \;=\; \{\tau_1, \tau_2, \tau_3\}\text{, where}$$

$$\tau_1 = [\;:[\text{x: dog(x), have(Peter, x)}] \Rightarrow [\text{y: gray(y), }\underline{\underline{\text{have(Peter, y), cat(y)}}}]]$$

$$\tau_2 = [\;:[\text{x, y: dog(x), have(Peter, x), }\underline{\underline{\text{have(Peter, y), cat(y)}}}] \Rightarrow [\;:\text{gray(y)}]]$$

$$\tau_3 = [\text{y: }\underline{\underline{\text{have(Peter, y), cat(y)}}}\text{, }[\text{x: dog(x), have(Peter, x)}] \Rightarrow [\;:\text{gray(y)}]]$$

Intuitively, the interpretation given by τ_3 (global accommodation) seems to be strictly preferred. This conforms to our intuition which interprets A by assuming that Peter has a cat and saying that it is gray in case Peter has a dog.

Another example is the following:

(16) A: If Peter has a cat, then his cat is gray

$$\text{sem}(A) \;=\; [\;:[\text{x: cat(x), have(Peter, x)}] \Rightarrow [\;:\text{gray(y)} \;/\; [\text{y: have(Peter, y), cat(y)}]]]$$

$$\text{Gen}(A) \;=\; \{\tau_1, \tau_2, \tau_3\}\text{, where}$$

$$\tau_1 = [\;:[\text{x: cat(x), have(Peter, x)}] \Rightarrow [\text{y: gray(y), }\underline{\underline{\text{have(Peter, y), cat(y)}}}]]$$

$$\tau_2 = [\;:[\text{x: }\underline{\text{cat(x), have(Peter, x)}}] \Rightarrow [\;:\text{gray(x)}]]$$

$$\tau_3 = [\text{y: }\underline{\underline{\text{have(Peter, y), cat(y)}}}\text{, }[\text{x: cat(x), have(Peter, x)}]\ \text{Y}\ [\;:\text{gray(y)}]]$$

In this case, the local projection (τ_1) and the global projection (τ_3) require accommodation. In contrast, the intermediate projection allows factoring, which is already realized in τ_2. (Bound material is indicated by single underlining). In example (16) the intuitively correct interpretation refers to the intermediate projection (τ_2).

In order to account for the intuitively correct interpretations of complex sentences that contain presupposition inducers, van der Sandt (1992) assumes that the projection process is restricted by general preferences. Geurts (1995) has reformulated and improved van der Sandt's account. His preferences are as follows:

(i) If a presupposition can both be bound or accommodated, there will in general be a preference for the first option, and
(ii) If a presupposition can be accommodated at two different sites, one of which is subordinate to the other, the higher site will, ceteris paribus, be preferred. (Geurts 1995: 27ff)

Moreover, Geurts provides a clear motivation for these preferences.

The rationale behind (i) is that hearers generally aim at interpretations that are maximally coherent, and (ii) is explained by the assumption that hearers tend to prefer the strongest interpretation that is consistent with what the speaker says (Geurts 1995: 28).[10]

My suggestion for an OT treatment of presupposition projection is simply to take the rationale behind Geurt's preferences more serious than the preferences themselves. Consequently, the following constraints can be formulated:

C1: *Avoid Accommodation* (*AvoidA*): It counts the number of discourse markers that are involved in accommodation.
C2: *Be Strong:* It evaluates pairs $\langle A, \tau \rangle$ with stronger outputs τ higher than pairs with weaker ones.

Their ranking is

R: *AvoidA* $\gg$ *BeStrong*

The first constraint prefers to bind presupposed material instead of accommodating it. Moreover, the present formulation of *AvoidA* gives a partial explanation for the preference for bridging and partial resolution over pure accommodation.[11] The notion of strength, on the other hand, is based on the entailment relation which is well defined within DRT (cf. Geurts 1995). As demonstrated in Blutner (1998), this notion can be refined by introducing a probabilistic measure. In any case, what is important is the fact that *BeStrong* is a graded constraint, not an absolute one. The ranking *AvoidA* $\gg$ *BeStrong* is necessary to validate van der Sandt's/Geurts' first preference.[12]

It is not difficult to see how interpretation optimality (I-principle) solves the selection task with regard to the examples given in (15) and (16). The respective OT tableaus are presented in (17) and (18) in a schematic form.

(17)

Ø		u > v	w > v
If p then q/r	☞ *AvoidA ^{u}BeStrong	*AvoidA ^{v}BeStrong	*AvoidA ^{w}BeStrong
	r, p ⇒ q (global)	(r ∧ p) ⇒ q (Interm.)	p ⇒ (q ∧ r) (local)

[10] In a footnote, Geurts tells us that this is true only as long as we ignore bridging. In the present paper, we are susceptible to this ignorance.

[11] By introducing probabilistic notions such as salience and cue validity the formulation of the constraint can be refined (perhaps along the lines outlined in Blutner 1998).

[12] I am convinced that this strict ranking system must be replaced by a cumulative constraint weighting system when it comes to considering the bulk of bridging phenomena.

In the first case all the possible outcomes (τ_1, τ_2, τ_3) violate the constraint *AvoidA* (with regard to the reference marker y). Consequently, *BeStrong* is the critical constraint. Because global accommodation gives the strongest outcome it wins the competition.

(18)

∅		u > v	w = v
If p then q/p	*AvoidA ^{u}BeStrong	☞ AvoidA ^{v}BeStrong	*AvoidA ^{w}BeStrong
	p, $p \Rightarrow q$ (global)	$p \Rightarrow q$ (Interm.)	$p \Rightarrow (q \wedge p)$ (local)[13]

In the second case, global and local projection give outcomes that violate the constraint *AvoidA*. In contrast, intermediate projection allows factoring and that is why it avoids accommodation. Because the constraint *AvoidA* ranks higher than the constraint *BeStrong*, intermediate projection is the winner.

Obviously, there is no necessary connection between how close the projection is to the main DRS and how strong the resulting interpretation is. A case in point where the two criteria diverge is given by the following example:

(19) a. Every German is proud of his car
b. Every German who owns a car is proud of it
c. Every German has a car and is proud of it

In (19a) global accommodation is excluded[14] and we have to select between intermediate and local accommodation only. Local accommodation refers to the stronger interpretation and intermediate accommodation refers to accommodation at the higher site. Consequently, if we take the criterion that prefers the higher site, then the interpretation of (19a) is identified with that of (19b). In contrast, the criterion that prefers the stronger interpretation identifies the interpretation of (19a) with that of (19c). Unfortunately, it is not easy to determine what the intuitively correct interpretation of (19a) is, since the proposition that *Germans have cars* is nearly tautological. Beaver (1994) gives an example where the judgement is easier. The following is a slightly simplified version.

[13] In this schematic formulation (ignoring reference markers) the intermediate and the local version seem to be logically equivalent, which is not really the case.

[14] The presupposition triggered by *his car* contains a reference marker that is bound by the quantifier and it would be free if the presupposition were accommodated globally (resulting in an improper DRS).

(20) a. ??Few of the team members can drive, but every team member will come to the match in her car.
b. Few of the team members can drive, but every team member who owns a car will come to the match in her car.
c. ?Few of the team members can drive, but every team member owns a car and will come to the match in her car

Intuitively the interpretation of (20a) is rather strange while (20b) is a perfectly acceptable sentence. According to Beaver (1994), this demonstrates that the van der Sandt/Geurts proposal must be wrong, since their criterion identifies the interpretation of (20a) with that of (20b). In contrast, the present OT proposal identifies the interpretation of (20a) with that of (20c), which I think is a much better choice.[15]

A further point is that we should explain why in many examples intermediate accommodation is clearly dominant, such as in the following:

(21) a. Birds lay eggs (preferred *female birds lay eggs*)
b. Most ships unload at night (preferred *most ships that unload do it at night*)

My feeling is that intermediate accommodation is partial in these cases and can outrank local accommodation, which is less partial.[16] The kind of partiality I have in mind is probabilistic in nature. A possible way to approach this phenomenon is by adopting an OT framework that is controlled by cue validity and other probabilistic factors (cf. Blutner (1998) for realizing such a framework using a Generator based on abduction). Further research seems necessary to clarify this point.

So far we have almost exclusively considered interpretation optimality (I-principle). Is it necessary to make use of the other way of optimization (Q-principle)? The answer is clearly affirmative. The point is that accommodation is not always possible although the I-principle demands it. Accommodation can be blocked. The following example by Asher & Lascarides (1998) gives a demonstration. Let us compare the two dialogues (22abc) and (22abd):

(22) a. A: Did you hear about John?
b. B: No, what?
c. A: He had an accident. A car hit him.
d. A: He had an accident. ??The car hit him.

[15] This is a somewhat unfair and roughly simplifying look on the van der Sandt/Geurts proposal. Geurts and van der Sandt (1999) demonstrate that with a little use of abstraction rules and propositional reference markers the data of Beaver (1994) can be handled. My point here is only to demonstrate that the problems can be resolved in a different way if we take the rationale behind the preferences more seriously than the preferences themselves.

[16] Note also the importance of stress and focus, especially in example (21b) (cf. Hendriks & de Hoop, to appear)

The van der Sandt/Geurts approach does not predict any difference between these two discourses and would find them both acceptable. But (22abd) is unacceptable, while (25abc) is acceptable. As a matter of fact the presupposition of *the car* cannot be accommodated in (22abd). With the help of the Q-principle this observation is easy to explain. Starting with a neutral context σ (neutral with regard to cars), the outcome of context change is the same for (22c) and for (22d). Consequently, the two sentences constitute simple expression alternatives. The difference is that in the second case but not in the first one accommodation is necessary to yield the output context. This makes the second case the more complex one and as such it is blocked by the simpler alternative (Q-principle).[17]

Zeevat (1999) formulated and substantiated the following theorem which generalizes a series of related facts. It can be proved in the very same way we have just sketched.

(23) A trigger for presuppositions does not accommodate iff any occurrence of it has a simple expression alternative that does not trigger.

Based on the availability of expression alternatives and the logical requirement of the presupposition proposed a fine-grained classification of presupposition triggers can be proposed. Even more interesting, an understanding of presupposition triggers like discourse particles, which are typically outside the scope of most standard theories becomes feasible (cf. Zeevat 1999).

The semantics and pragmatics of focus provides a further challenge for applying the present ideas. Adding only one new constraint, *Avoid Focus*, which is ranked lower than *Avoid Accommodation*, it is a simple exercise to demonstrate that Schwarzschild's deaccenting theory of congruence (Schwarzschild 1999) is a natural consequence of the present ideas, crucially making use of the Q-principle.

In the first part of this paper I have outlined some theoretical reasons that recommend the weak version of bidirectional OT. From an empirical point of view it is not trivial to find data where the weak version is clearly

[17] Bart Geurts (p.c.) argues that the discourse (25d) is unacceptable because the proposition made by the second part is rather uninformative (supposed appropriate bridging). Though this idea is interesting it cannot be the whole story. In particular, the idea cannot explain the contrast between the following examples:

c′. He had a bike accident. A car hit him seriously.
d′. He had a bike accident. ?The car hit him seriously.

Furthermore, the contrast does not disappear when dropping the material that according to Geurts can trigger bridging:

c″. A car hit him (seriously).
d″. ?The car hit him (seriously).

preferred over its strong counterpart. The investigation of phenomena where Q-based effects (blocking) interact with I-based effects (interpretational preferences) may be an opportunity to make the comparison conceivable. As a first step in this direction, Jäger & Blutner (to appear) investigated the interaction between polysemy and focus. Dealing with the German adverb of repetition 'wieder' (again), the specific linguistic puzzle that was envisaged concerned the selection of the repetitive vs. the restitutive readings, depending on focus and scrambling. The results appeared to favour the weak version of bidirectional OT. It seems important to me to pursue the problem of discriminating between the weak and the strong version in depth.

Acknowledgements

This paper is dedicated to Manfred Bierwisch on the occasion of his 70th birthday. This work was supported by the Deutsche Forschungsgemeindschaft (DFG). Parts of this paper were first presented on a DIP colloquium in Amsterdam. My special thanks go to Henk Zeevat and Helen de Hoop who have encouraged me to pursue this line of research and gave valuable impulses and stimulation. Furthermore, I have to thank Anton Benz, Manfred Bierwisch, Paul David Doherty, Werner Frey, Bart Geurts, Gerhard Jäger, Paul Law, Klaus Robering, Paul Smolensky, and Rob van der Sandt. I am grateful to two anonymous referees for their very helpful comments.

REINHARD BLUTNER
Humboldt University, Berlin
Prenzlauer Promenade 149–152
D-13189 Berlin
Germany
blutner@web.de
http://www2.rz.hu-berlin.de/asg/blutner/

Received: 05.04.00
Final version received: 25.07.00

REFERENCES

Asher, Nicholas & Lascarides, Alex (1998), 'The Semantics and Pragmatics of Presupposition', *Journal of Semantics*, 15, 239–99.

Beaver, David (1994), 'Accommodating Topics', in Peter Bosch & Rob van der Sandt (eds), *Focus and Natural Language Processing. Volume 3: Discourse*, IBM, Heidelberg, 439–48.

Atlas, Jay David & Levinson, Stephen C. (1981), 'It-clefts, informativeness and logical form', in Peter Cole (ed.), *Radical Pragmatics*, Academic Press, New York, 1–61.

Bierwisch, Manfred (1983), 'Semantische Einheiten und konzeptuelle Repräsentation lexikalischer Einheiten', in *Untersuchungen zur Semantik*, Akademie-Verlag, Berlin, 61–99.

Bierwisch, Manfred (1996), 'Lexical information from a minimalistic point', in Chris Wilder (ed.), *The Role of Economy Principles in Linguistic Theory*, Akademie-Verlag, Berlin. 227–66.

Blutner, Reinhard (1998), 'Lexical pragmatics', *Journal of Semantics*, 15, 115–62.

Boersma, Paul (1998), *Functional Phonology*, Holland Academic Graphics, The Hague.

Bresnan, Joan (to appear), 'Explaining morphosyntactic competition', in Mark Baltin & Chris Collins (eds), *Handbook of Contemporary Syntactic Theory*, Blackwell, Oxford.

Burzio, Luigi (1989), 'On the non-existence of disjoint reference principles', *Rivista di Grammatica Generativa*, 14, 3–27.

Burzio, Luigi (1998), 'Anaphora and soft constraints', in Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis, & David Pesetsky (eds), *Optimality and Competition in Syntax*, MIT Press, Cambridge, MA, 93–113.

Carston, Robyn (1998), 'The semantics/pragmatics distinction: a view from relevance theory', *UCL Working Papers in Linguistics*, 10, 1–30.

Choi, Hye-Won (1996), 'Optimizing Structure in Context', Ph.D. dissertation, Stanford University.

Copestake, Ann & Briscoe, Ted (1995), 'Semi-productive polysemy and sense extension', *Journal of Semantics*, 12, 15–67.

Deemter, Kees van & Peters, Stanley (eds) (1996), *Semantic Ambiguity and Underspecification*, CSLI Publications, Stanford, CA.

Dekker, Paul & Rooy, Robert van (this volume), 'Optimality theory and game theory: some parallels'.

Fanselow, Gisbert, Schlesewsky, Matthias Cavar, Damir & Kliegl, Reinhold (1999), 'Optimal parsing'. MS, University of Potsdam.

Geurts, Bart (1995), 'Presupposing', Ph.D. dissertation, University of Osnabrück.

Geurts, Bart (1999), 'Presuppositions and Pronouns', Elsevier, Oxford.

Geurts, Bart & Sandt, Robert A. van der (1999), 'Domain restriction', in Peter Bosch & Robert A. van der Sandt (eds), *Focus: Linguistic, Cognitive and Computational Perspectives*, Cambridge University Press, Cambridge.

Gibson, Edward & Broihier, Kevin (1998), 'Optimality theory and human sentence processing', in Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis, & David Pesetsky (eds), *Optimality and Competition in Syntax*. MIT Press, Cambridge, MA, 157–91.

Groenendijk, Jeroen & Stokhof, Martin (1991), 'Dynamic predicate logic', *Linguistics and Philosophy*, 14, 39–100.

Haspelmath, Martin (1999), 'Optimality and diachronic adaptation', *Zeitschrift für Sprachwissenschaft*, 18, 180–205.

Heim, Irene (1982), 'The semantics of definite and indefinite noun phrases', Ph.D. thesis, University of Massachusetts, Amherst.

Hendriks, Petra & Hoop, Helen de (to appear), 'Optimality theoretic semantics', *Linguistics and Philosophy*.

Hoop, Helen de (2000), 'Optimal scrambling and interpretation', in H. Bennis, M. Everaert, & E. Reuland (eds), *Interface Strategies*, KNAW, Amsterdam, 153–168.

Hoop, de Helen & Swart, Henriëtte de (1998), 'Temporal adjunct clauses in optimality theory', MS, OTS Utrecht.

Horn, Laurence 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*, Georgetown University Press, Washington, 11–42.

Itô, Junko, Mester, Armin & Padgett, Laye (1995), 'Underspecification in optimality theory', *Linguistic Inquiry*, 26, 571–613.

Jäger, Gerhard (1999), 'Optimal syntax and optimal semantics', handout for talk at DIP-colloquium. Available from http://www.zas.gwz-berlin.de/mitarb/homepage/jaeger/

Jäger, Gerhard & Blutner, Reinhard (to appear), 'Against lexical decomposition in syntax', in *Proceedings of IATL 15*, University of Haifa.

Jakobson, Roman (1941/1968), *Child Language, Aphasia and Phonological Universals*, Mouton, The Hague.

Kadmon, Nirit (1990), 'Uniqueness', *Linguistics and Philosophy*, 13, 273–324.

Kager, René (1999), *Optimality Theory*, Cambridge University Press, Cambridge.

Kamp, Hans (1981), 'A theory of truth and semantic representation', in Jeroen Groenendijk *et al.* (eds), *Formal Methods in the Study of Language*, Mathematisch Centrum, Amsterdam.

Kamp, Hans & Reyle, Uwe (1993), *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Kluwer Academic Publishers, Dordrecht.

Karttunen, Lauri (1998), 'The proper treatment of optimality in computational phonology', Xerox Research Centre Europe manuscript. ROA-258-0498, Rutgers Optimality Archive, http://ruccs.rutgers.edu/roa.html.

Kiparsky, Paul (1982), 'Word-formation and the lexicon' in F. Ingeman (ed.), *Proceedings of the 1982 Mid-America Linguistic Conference.*

Kuhn, Jonas (2000), 'Generation and parsing in optimality theoretic syntax–issues in the formalization of OT-LFG', to appear in Peter Sells (ed.), *Formal and Empirical Issues in Optimality-theoretic Syntax*, CSLI Publications, Stanford, CA.

Lee, Hanjung (2000), 'Markedness and word order freezing', to appear in Peter Sells (ed.), *Formal and Empirical Issues in Optimality-theoretic Syntax*, CSLI Publications, Stanford, CA.

Levinson, Stephen C. (1983), *Pragmatics*, Cambridge University Press, Cambridge.

Levinson, Stephen C. (1987), 'Pragmatics and the grammar of anaphora', *Journal of Linguistics*, 23, 379–434.

McCawley, James D. (1978), 'Conversational implicature and the lexicon', in Peter Cole (ed.), *Syntax and Semantics 9: Pragmatics*, Academic Press, New York, 245–59.

Nunberg, G. & Zaenen, A. (1992), 'Systematic polysemy in lexicology and lexicography', in K. Varantola, H. Tommola, T. Salmi-Tolonen, & J. Schopp (eds), *Euralex II*, Tampere, Finland.

Pesetsky, David (1997), 'Optimality Theory and syntax: movement and pronunciation', in Diana Archangeli & D. Terence Langendoen (eds), *Optimality Theory: An Overview*, Blackwell, Oxford, 134–70.

Prince, Alan & Smolensky, Paul (1993), 'Optimality theory: constraint interaction in generative grammar., MS, Rudgers University, New Brunswick, NJ and University of Colorado, Boulder (to appear, MIT Press, Cambridge, MA).

Reyle, Uwe (1993), 'Dealing with ambiguities by underspecification: construction, representation and deduction', *Journal of Semantics*, 10, 123–79.

Sandt, van der & Robert, A. (1992), 'Presupposition projection as anaphora resolution', *Journal of Semantics*, 9, 333–77.

Schwarzschild, Roger (1999), 'GIVENness, AvoidF and other constraints of the placement of accent', *Natural Language Semantics*, 13, 87–138.

Smolensky, Paul (1986), 'Information processing in dynamical systems: foundation of harmony theory', in David E. Rumelhart & James L McClelland (eds), *Explorations in the Microstructure of Cognition*, MIT Press, Cambridge, MA, 194–281.

Smolensky, Paul (1996), 'On the comprehension/production dilemma in child language', *Linguistic Inquiry*, 27, 720–31.

Speas, Margaret (1997), 'Optimality theory and syntax: null pronouns and control', in Diana Archangeli & D. Terence Langendoen (eds), *Optimality Theory: An Overview*, Blackwell, Oxford, 134–70.

Tesar, Bruce & Smolensky, Paul (2000), *Learnability in Optimality Theory*, MIT Press, Cambridge, MA.

Williams, Edwin (1997), 'Blocking and anaphora', *Linguistic Inquiry*, 28, 577–628.

Wilson, Colin (1998), 'Bidirectional optimization and the theory of anaphora', MS, Johns Hopkin University, to appear in Jane Grimshaw, Géraldine Legendre, & Sten Vikner (eds), *Optimality Theoretic Syntax*, MIT Press, Cambridge, MA.

Wurzel, Wolfgang U. (1998), 'On markedness', *Theoretical Linguistics*, 24, 53–71.

Zeevat, Henk (1999), 'Explaining presupposition triggers', MS AC99, University of Amsterdam, available from http://www.hum.uva.nl/computerlinguistiek/henk/

Zeevat, Henk (this volume), 'Semantics in optimality theory'.

Zipf, George K. (1949), *Human Behavior and the Principle of Least Effort*, Addison-Wesley, Cambridge.