Some Experimental Aspects of Optimality-Theoretic Pragmatics

Reinhard Blutner
Universiteit van Amsterdam

Abstract

The article has three main concerns: (i) it gives a concise introduction into optimality-theoretic pragmatics; (ii) it discusses the relation to alternative accounts (relevance theory and Levinson's theory of presumptive meanings); (iii) it reviews recent findings concerning the psychological reality of optimality-theoretic pragmatics and its central part concept – bidirectional optimization.

　　A present challenge is to close the gap between experimental pragmatics and neo-Gricean theories of pragmatics. I claim that OT pragmatics helps to overcome this gap, in particular in connection with the discussion of asymmetries between natural language comprehension and production. The theoretical debate will be concentrated on two different ways of interpreting bidirection: (a) bidirectional optimization as a psychologically realistic online mechanism; (b) bidirectional optimization as an offline phenomenon of fossilizing optimal form-meaning pairs. It will be argued that neither of these extreme views fits completely with the empirical data when taken per se.

1 Introduction

Recent approaches to experimental pragmatics (see Noveck & Sperber, 2005) mainly concentrated on the investigation of scalar implicatures. Characteristically, the interpretive perspective (hearer's view) is taken in this research. A theoretical main issue is to decide between two rivaling theories: Sperber & Wilson's (1986) Relevance Theory (RT) and Levinson's (2000) theory of presumptive meanings or generalized conversational implicatures (GCIs). Levinson claims that GCIs are calculated automatically – i.e. without demanding much processing resources. In contrast, RT argues that the calculation is controlled and is strongly influenced by the available processing resources. Neo-Griceans (Atlas & Levinson, 1981; Horn, 1984; Blutner, 1998; e.g. Atlas, 2005; Horn, 2005) are normally ignored in this research. A defense for this pretermission is that neo-Gricean theories are normative theories that do not directly make predictions about processing. Unfortunately, this argument exaggerates the philosophical issue of distinguishing between the normative and the naturalistic realm. Surely, these two aspects of understanding human actions can be clearly separated from each other. However, that does not mean they predict different action patterns in most cases. The idea of a rational world isn't so irrational to be excluded in ordinary affairs. Evolutionary game theory has presented us with many examples demonstrating that the reasonable is naturally arising (Axelrod, 1984). In other words, though there is a philosophical gap between Gricean pragmatics as a normative theory and experimental pragmatics as a scientific, explanatory theory of natural language interpretation, there is no deep empirical conflict between interpretation oriented pragmatics and speaker ethics. It seems the speaker better be cooperative (or pretend to be cooperative) if she wants to use language to bring about effects in hearers.

　　The aim of this article is to close the gap between experimental pragmatics and neo-Gricean theories of pragmatics. The version of neo-Gricean pragmatics I will consider here is called optimality-theoretic (OT) pragmatics. While the automatic/controlled issue of processing has dominated the recent theoretical debate, OT pragmatics will raise several additional issues. One new issue concerns the asymmetries between comprehension and

production. How to explain the experimentally observed asymmetries and what is their status in theories of language acquisition? Seeing both comprehension and production as different optimization processes, a further research topic concerns the question of how the two optimization processes are integrated with each other (bidirectional optimization). That relates to the psychological reality of bidirectional optimization in the domain of pragmatics. Another new concerns the nature of conventionalization (or fossilization) in pragmatics – an issue that deserves much more attention both theoretically and empirically.

The following quotation from Noveck & Sperber (2005) fully applies to the raised new pragmatic issues.

> Properly devised experimental evidence can be highly pertinent to the discussion of pragmatic issues, and pragmatics might greatly benefit from becoming familiar with relevant experimental work and from contributing to it. (Noveck & Sperber 2005, p. 210)

Without careful experimental research linguistic pragmatics cannot really mature and will remain in a phase of rampant speculation and questionable research habits.

Optimality theory (OT) will be used in this article both in the *broad sense* of a general methodology dealing with resolving conflicting constraints by using universal optimization procedures and in the *narrower sense* of developing an explicit model that concern the essentials of neo-Gricean pragmatics.

In the following section I give a concise introduction to the basic conceptions of OT pragmatics, and I explains the idea of unidirectional and bidirectional optimization. In section 3 the three main views conforming to a naturalistic pragmatics are discussed: (a) RT, (b) Levinson's (2000) theory of presumptive meanings, and (c) the neo-Gricean approach. I show how the idea of optimal interpretation can be used to restructure the core ideas of these three different approaches. Section 4 explains the idea of fossilization. It is pointed out how the general setting of cultural evolution can help to make this idea precise. Further, a series of important theoretical problems is raised - mainly concerning the distribution of labor between online processes (optimization procedures) and offline processing (fossilization processes). In section 5 I discuss a series of experimental findings and come to a preliminary conclusion about the relationship between online processes and fossilization phenomena. Section 6 draws some general conclusions relating to a deeper understanding of the idea of naturalization and (cultural) embodiment in the context of natural language interpretation.

2 OT Pragmatics

Optimality Theory is an approach to embodied cognition in the sense of Anderson (2003). It combines the advantages of constraint-based, symbolic models with the advantages of neuronal network models of cognition (cf. Smolensky & Legendre, 2006). Further, it has the capacity to model cultural evolution and to solve the symbolic grounding problem (e.g. Harnad, 1990; Boersma & Hayes, 2001). In the study of natural language, OT was successfully applied to the main linguistic disciplines of phonology, morphology and syntax, and also to the explanation of natural language acquisition and other performance traits. OT pragmatics is an application of this integrated approach to the domain of Gricean pragmatics. It has its origin in the attempt to explain certain phenomena of lexical pragmatics (Blutner, 1998) and is inspired by the optimal interpretation approach proposed by Hendriks & de Hoop (2001) and Zeevat's (1999) theory of explaining presupposition triggers.

In the broad view of OT, this framework can be seen as a general framework that systematizes the use of optimization methods in linguistics.[1] One component of OT is a list of tendencies that hold for observable properties of a language. These tendencies take the form of violable constraints. Because the constraints usually express very general statements, they can be in conflict. Conflicts among constraints are resolved because the constraints differ in strength. Minimal violations of the constraints (taking their strength into account) define optimal conflict resolutions.

Standardly, OT specifies the relation between an input and an output. This relation is mediated by two formal mechanisms, **GEN** and **EVAL**. GEN (for Generator) creates possible output candidates on the basis of a given input. EVAL (for Evaluator) uses the particular constraint ranking of the universal set of constraints **CON** to select the best candidate for a given input from among the candidate set produced by GEN. In phonology and syntax, the input to this process of optimization is an underlying linguistic representation. The output is the (surface) form as it is expressed. Hence, what is normally used in phonology and syntax is unidirectional optimization. Obviously, the point of view of the speaker is taken. This contrasts with OT semantics where the view of the hearer is taken (de Hoop & de Swart, 2000; Hendriks & de Hoop, 2001).

Bidirectional optimization (Blutner, 1998, 2000) integrates the speaker and the hearer perspective into a simultaneous optimization procedure. In pragmatics, this bidirectional view is motivated by a reduction of Grice's maxims of conversation to two principles: the R-principle, which can be seen as the force of unification minimizing the Speaker's effort, and the Q-principle, which can be seen as the force of diversification minimizing the Auditor's effort. In the next section we will explain the relationship between the neo-Gricean view and OT pragmatics in a bit more detail. In this section I will introduce the heart of OT pragmatics – bidirectional optimization. Further, I will give a very schematic example in order to illustrate some characteristics of bidirectional OT.

In the so-called strong version of bidirectional OT, the following clause applies:

(1) Strong Bidirection: A form-interpretation pair $<f, m>$ is considered to be (strongly) optimal iff
   a. Interpretive Optimization: no other pair $<f, m'>$ can be generated that satisfies the constraints better than $<f, m>$ and
   b. Expressive Optimization: no other pair $<f', m>$ can be generated that satisfies the constraints better than $<f, m>$.

The idea behind interpretive optimization is to select the most coherent interpretation. What is meant by coherence has to be expressed by particular OT constraints, such as formulated by Zeevat (2007), for instance. Hence, the principle of interpretive optimization is a very abstract one which has to be supplemented by a system of ranked constraints in order to constitute a system that is able to express something like Horn's R-principle. The simultaneous use of expressive optimization can be seen as similar to the role of Horn's Q-principle - it acts as a blocking mechanism which blocks all the outputs which can be expressed more economically by an alternative linguistic input. Again, what counts as more economical has to be expressed by the system of constraints. This formulation makes it quite clear that the neo-Gricean framework can be conceived of as a bidirectional optimality framework which integrates the speaker and the hearer perspective. Whereas the R-principle compares different possible interpretations for the same syntactic expression, the Q-principle compares different possible syntactic expressions that the speaker could have used to communicate the same meaning.

---

[1] A recent overview is given in Smolensky & Legendre (2006). For OT pragmatics the reader is referred to Blutner & Zeevat (2004) and Blutner, de Hoop & Hendriks (2005).

The following example gives a simple illustration of how the theory works and how the required ingredients look like. Assume we have two forms f_1 and f_2 which are semantically equivalent. This means that **GEN** associates the same interpretations with them, say m_1 and m_2. We stipulate that the form f_1 is less complex (less marked) than the form f_2 and that the interpretation m_1 is less complex (less marked) than the interpretation m_2. This is expressed by two markedness constraints: F for forms and M for interpretations – F prefers f_1 over f_2 and M prefers m_1 over m_2. This is indicated by the two leftmost constraints in table 1.

	F	M	F→M	*F→*M	F→*M	F*→M
$<f_1, m_1>$					*	
$<f_1, m_2>$		*	*			
$<f_2, m_1>$	*			*		
$<f_2, m_2>$	*	*				*

Table 1: Markedness and linking constraints in a 2-forms × 2-interpretations design

Besides the markedness constraints, four so-called linking constraints can be formulated. There are precisely four independent linking constraints in the present example. The linking constraint F→M says that simple (unmarked) forms express simple interpretations. The constraint *F→*M says that complex forms express complex interpretations. The two remaining linking constraints express the opposite restrictions. In the present case linking constraints can be seen as lexical stipulations that fix a form-interpretation relation in an instance-based way.[2]

From the differences of markedness given by the constraints F and M the ordering relation between form-meaning pairs can be derived as shown in Figure 1. The preferences are indicated by arrows in a two-dimensional diagram. Such diagrams give an intuitive visualization for the optimal pairs of (strong) bidirectional OT: they are simply the meeting points of horizontal and vertical arrows.[3] The optimal pairs are marked with the symbol ☝ in the diagram.

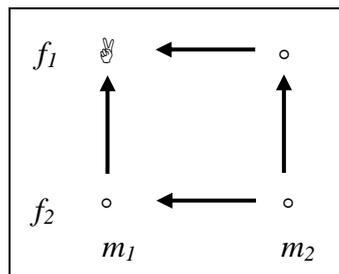
Figure 1: Diagram to illustrate strong bidirection

The scenario just mentioned 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. This means that not all the interpretations of a form must be blocked if another form exists. McCawley (1978) collects a number of examples demonstrating the phenomenon of *partial blocking*. For example, he observes that the distribution of productive causatives (in English, Japanese, German, and other languages) is restricted by the existence of a corresponding lexical causative. Whereas lexical causatives (e.g. (2a)) 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, (2b) could have been used appropriately when Black Bart caused the sheriff's gun to backfire by stuffing it with cotton.

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

To make things concrete we can take f_1 to be the lexical causative form (2a), f_2 the periphrastic form (2b), m_1 direct (stereotypic) causation and m_2 indirect causation.

Typical cases of partial blocking are found in morphology, syntax and semantics. 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) terms *"the division of pragmatic labour"*.

There are two ways of avoiding total blocking within the bidirectional OT framework and to describe Horn's division of pragmatic labour. The first possibility makes use of linking constraints and fits the intended form-interpretation relation by stipulating the appropriate ranking of the constraints such that partial blocking comes out. Let's assume that the two linking constraints F→M and *F→*M are higher ranked than the rest of the constraints. This can be depictured as in Figure 2a. Hence, strong bidirection can be taken as describing Horn's division of pragmatic labour when the appropriate linking constraints are dominant.
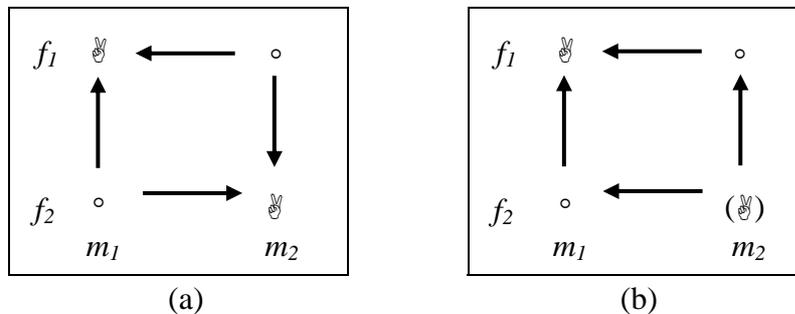

Figure 2: Two ways of describing Horn's division of pragmatic labour: (a) by assuming the two dominant linking constraints F→M and *F→*M and strong bidirection; (b) by assuming the markedness constraints F and M and weak bidirection

The second possibility is to weaken the notion of (strong) optimality in a way that allows us to derive Horn's division of pragmatic labour by means of the *evaluation procedure* and without stipulating particular linking constraints. Blutner (2000) proposes a *weak* version of two-dimensional OT, according to which the two dimensions of optimization are mutually related. The definition for super-optimality is as follows:

(3) Weak Bidirection (Super-Optimality): A form-interpretation pair <*f*, *m*> is considered to be super-optimal iff

a. Interpretive Optimization: no other *super-optimal* pair $\langle f, m'\rangle$ can be generated that satisfies the constraints better than $\langle f, m\rangle$;

b. Expressive Optimization: no other *super-optimal* pair $\langle f', m\rangle$ can be generated that satisfies the constraints better than $\langle f, m\rangle$.

This formulation looks like a circular definition, but Jäger (2002) has shown that this is a sound *recursive* definition under very general conditions (well-foundedness of the ordering relation). The important difference between the weak and strong notions of optimality is that the weak one accepts super-optimal form-meaning pairs that would not be optimal according to the strong version. It typically allows marked expressions to have an optimal interpretation, although both the expression and the situations they describe have a more efficient counterpart.

Figure 2b shows that the *weak* version of bidirection can explain the effects of partial blocking without the stipulation of extra linking constraints; especially it can explain why the marked form f_2 gets the marked interpretation m_2. This is a consequence of the *recursion* implemented in weak bidirection: the pairs $\langle f_1, m_2\rangle$ and $\langle f_2, m_1\rangle$ are *not* super-optimal. Hence, they cannot block the pair $\langle f_2, m_2\rangle$ and it comes out as a new super-optimal pair. In this way, the weak version accounts for Horn's pattern of *the division of pragmatic labour*.

The two parts of Figure 2 describe the same set of solution pairs but the calculation of the solutions is completely different in the two cases. In the first case unidirectional optimization (either hearer or speaker perspective) is sufficient to calculate the solution pairs. It is plausible to assume that this kind of OT systems can be used to construct cognitively realistic models of online, incremental interpretation (cf. Blutner, 2006, 2007). The second case – using the recursion of weak bidirection (super-optimality) – has a completely different status. Because of its strictly non-local nature the proposed algorithms that calculate the super-optimal solutions seem not to fit the requirements of a psychologically realistic model of online, incremental interpretation. Rather, the proper understanding of weak bidirection relates best to an off-line mechanism of bidirectional learning and fossilization. The second half of section 4 explains this idea in detail.

3 Three Variants of Gricean Pragmatics

The *naturalization of pragmatics* refers to a research program that aims to provide a cognitively realistic picture of utterance interpretation and production. Hence, the proponents of this program such as relevance theorists take the stance of seeing natural language interpretation as a cognitive phenomenon and thus considering the basic principles of communication as a consequence of the nature of human cognition. A prerequisite of this program deals with the levels of cognitive representations and the boundary between semantics and pragmatics. As pointed out by Atlas (2005) and Jaszczolt (to appear), there is a strong tendency among current researchers to follow the tradition of radical pragmatics and to accept the following three claims:

▪ There is a level of logical form or semantic representation. The representations of this level do not necessarily provide truth conditions. Rather, they underspecify truth-conditional content in a number of ways.

▪ There is a mechanism of enriching underspecified representations.[4] The result of enriching is propositional content. It expresses the utterance meaning of the relevant expression.

---

[4] Sometimes this mechanism is called development of logical form.

- There is a level of implicatures proper, understood as separate thoughts implied by the utterance. It is implicit propositional content that can be inferred from the explicit content mentioned in 2.

Obviously, the consensus is about rejecting the Gricean doctrine of literal meaning (logical form conforms to literal meaning), accepting the role of underspecification (logical forms are underspecified with regard to the expressed semantic content) and acknowledging that implicature is a graded category (some implicatures are closer to LF than others). Obviously, this view sharply contrasts with the paradigm of generative semantics – a view that tries to ground pragmatic phenomena by using particular syntactic stipulations. There are still representatives of this view in linguistics. For a critical discussion the reader is referred to Russell (2006).

Before we discuss three variations on Grice and the naturalization of conversational implicatures it is useful first to introduce the distinction between *global* and *local* approaches to conversational implicatures (cf. Chierchia, 2004). According to the global (Neo-Gricean) view one first computes the (plain) meaning of the sentences; then, taking into account the relevant alternatives, one strengthens that meaning by adding in the implicature.' (Chierchia 2004: 42). This contrasts with the local view, which first introduces pragmatic assumptions locally and then projects them upwards in a strictly compositional way where certain filter conditions apply. Representatives of the global view are Atlas & Levinson (1981), Gazdar (1979), Horn (1984), Soames (1982), Krifka (1995), Blutner (1998), Sauerland (2004), and Sæbø (2004); the local view is taken by Chierchia (2004), Levinson (2000), and relevance theory (Sperber & Wilson, 1986; e.g. Carston, 2002).

Usually, the globalists argue against the local view and the localists against the global view. I have argued at another occasion (Blutner, 2006, 2007) that proper variants of both views are justified if a different status is assigned to the two views: global theories provide the standards of rational discourse and correspond to a diachronic, evolutionary scenario; local theories account for the shape of actual, online processing including the peculiarities of incremental interpretation. In this way, it can be claimed that seemingly conflicting approaches such as RT and the neo-Gricean approach are much closer related than the proponents (and opponents) of these theories expected.

I will take the broad view of OT in this section, seeing OT as a general scheme that can be used for expressing many different and possibly diverging views. For instance, it is possible to give optimality-theoretic reconstructions of a speaker-oriented normative pragmatics like the one developed by Grice. It is also possible to reconstruct hearer-oriented naturalistic pragmatics as in RT (cf. Hendriks & de Hoop, 2001; Zeevat, 2007). These systems are important for online synchronic accounts of speaking and interpretation. Interestingly, the main aspects of Levinson's (2000) theory of presumptive meanings can also be mimicked in a closely related way. Perhaps more surprisingly, the reconstruction of the neo-Gricean systems of Atlas & Levinson (1981) and Horn (1984) is possible but crucially deviates from the reconstruction of Levinson (2000). Similar to the reconstruction of RT, the reconstruction of Levinson (2000) requires a unidirectional optimization mechanism. This contrasts with the restoration of the neo-Gricean systems which requires two opposing optimization principles (corresponding to the Q- and the I/R-principle). The integration of two simultaneous optimization procedures (speaker and the hearer direction) automatically leads to a bidirectional OT pragmatics.

3.1 Relevance Theory

RT assumes the representational/computational view of the mind, and, on this basis, gives a naturalization of pragmatics adopting Jerry Fodor's language of thought hypothesis (Fodor, 1975). The central thesis of RT is the communicative principle of relevance, according to which utterances convey a presumption of their own optimal relevance. In other words, any given utterance can be presumed:

(i)   to be at least relevant enough to warrant the addressee's processing effort

(ii)   to be the most relevant one compatible with the speaker's current state of knowledge and her personal preferences and goals.

From these two assumptions relevance theorists derive the following general procedure that the cognitive system follows in comprehending an utterance (cf. Sperber, Cara, & Girotto, 1995: 95): (a) test possible interpretations in their order of accessibility, (b) stop once the expectation of (optimal) relevance is satisfied (i.e. a certain context-dependent threshold value of relevance is reached). The procedure makes sure that the desired effect (a certain value of relevance) is reached with the minimal cognitive effort.

Obviously, RT conforms to the localist approach and can be formulated in terms of unidirectional optimization. Let's stipulate a constraint EFFECT for describing the desired effect (a certain value of relevance) and a constraint EFFORT for describing the cognitive effort. Then the stipulation EFFECT >> EFFORT makes sure that the desired effect is reached with the minimal cognitive effort. Obviously, there are many questions left concerning the concrete content of the constraints EFFECT and EFFORT, the RT literature contains a number of specifications. These specifications typically have the character of linking constraints. It might be interesting to investigate recent OT models of pragmatics in the light of the general structure of RT – a task that goes beyond what can be done in the present paper.[5]

Another important aspect concerns pragmatic acceptability. The RT account of pragmatic (un)acceptability is worked out in connection with bridging phenomena (Wilson & Matsui, 1998). In RT, "unacceptability can result from (a) inadequate effects or (b) gratuitous effort" (Wilson & Matsui, 1998: 19). That means there have to be thresholds for (a) effects and (b) effort, and when these thresholds are reached unacceptability results. This kind of argumentation is not possible within an OT approach because in OT the absolute strength of constraint violation is not of importance. What counts is the comparison with other expressions that lead to the same interpretation and the possibility of blocking an interpretation by a cheaper expression alternatives (e.g. Blutner, 1998). It's an open issue to test the empirical consequences of this view.

---

[5] For instance, Zeevat (2002; 2007) reconstructed the presupposition theory as formulated by Van der Sandt (1992) and Heim (1983). In both these theories there are two preferences. The first one prefers identifying the induced presupposition in the context of the utterance (resolution), the second one prefers the addition of the presupposition to the global context (Heim) or to the highest accessible context where that is possible (Van der Sandt). Zeevat's reconstruction makes use of the following constraints that are used in finding an optimal interpretation: FAITH > CONSISTENCE > DO NOT ACCOMMODATE > STRENGTH. CONSISTENCE demands that there is no conflict of the current utterance with what is known already, FAITH asks for the presence of the presupposed information at an accessible position, DO NOT ACCOMMODATE forbids the addition of the presupposed information and STRENGTH forbids interpretations if there are informationally stronger ones available. Obviously, FAITH and DO NOT ACCOMMODATE are constraints of type EFFORT while CONSISTENCE and STRENGTH are realizations of Type EFFECT. This violates the RT scheme EFFECT > EFFORT saying that all EFFECT-constraints outrank all EFFORT-constraint.

3.2 Presumptive Meanings

Levinson's (2000) theory of presumptive meanings is a chameleon that in a certain sense adapts general assumptions of RT and in another sense crucially conflicts with RT, for instance in assuming more than one basic principle (*maxim*) for formulating the interpretational mechanism. In short, these are the general assumptions:

(i) Differing from both RT and the standard neo-Gricean view, Levinson assumes *three* levels of meaning corresponding to coded meaning, utterance-type meaning and utterance-token meaning

(ii) Coded meanings constitute what is normally called semantics, i.e. a convention of a language at a given time. In the present context it relates to the association of natural language forms with underspecified representations of meaning.

(iii) Utterance-type meanings are in correspondence with Grice' generalized conversational implicatures. More precisely, they are *fossilized* (or crystallized) invited inferences associated with certain lexemes or particular constructions. As a matter of preferred interpretation, they are calculated by a particular default mechanism which is relatively context-independent. It's in their nature that they may be cancelled. Basically, there are three such defaults or heuristics:

- Q-heuristic: *What isn't said is not the case*
- I-heuristic: *What is expressed simply is stereotypically exemplified*
- M-heuristic: *What's said in an abnormal way isn't normal* [6]

(iv) Utterance token meanings are invited inferences that have not been fossilized into commonly used implicatures. They are highly context-dependent and as such they may be based in encyclopaedic knowledge or in situation-specific information.

(v) In contrast to utterance token meanings, which are calculated in a global manner, utterance type meanings (= presumptive meanings) are local, i.e. they arise at the point at which they are triggered (for instance, the word *some* triggers the default interpretation NOT ALL via the Q-heuristics). The feature of local pragmatics is essential to artificial intelligence pragmatics (e.g. Hobbs & Martin, 1987) and likewise to RT.

Presumptive meanings are identical with utterance type meanings and based on generalized conversational implicatures (GCIs). The mechanics of GCIs is very useful for understanding natural language interpretation, especially for explaining the predominantly incremental character of utterance comprehension. Like RT Levinson's (2000) theory of presumptive meanings conforms to the localist approach and can be formulated in terms of unidirectional optimization. The defaults Q, I and M can be formulated as OT constraints. In case of two forms and two meanings (as considered in section 2; see table 1), the I-heuristics conforms to the linking constraint F→M saying that simple (unmarked) forms express simple interpretations. Similarly, the M-heuristics is expressed by the constraint *F→*M saying that complex forms express complex interpretations. With only two forms and two meanings, the substance of the Q-heuristics is not really different from that of the M-constraint. To get an independent extensional expression of the Q-heuristics there needs to be more than two forms and two meanings.

An advantage of the present OT approach is its general grounding in embodied cognition in general and in connectionist network research in particular. Using this methodology, we

---

[6] Levinson's M-heuristics should not be confused with the markedness constraint M introduced in Table 1.

can achieve a clarification concerning the nature of *automatic* vs. *controlled processing*. *Automatic processing* is highly parallel, but limited in power. *Controlled processing* has powerful operations, but is limited in capacity (cf. Schneider & Shiffrin, 1977). The distinction between these two types of processing has been used, for example, for modelling lexical access, visual perception, problem solving and parsing strategies in natural language processing. As demonstrated elsewhere (Blutner, 2004; Smolensky & Legendre, 2006), in cases where the constraint network is built with exponential weights[7] the energy landscape is structured like high mountains with an obvious path to the valley of the global energy minimum. Moreover, the stochastic solution algorithm yields robust results given some fluctuation of the parameters corresponding to attention.[8] Hence we have the tentative characterization of automatic processing. In the other case, however, the energy landscape is flat and there are many walls of comparable height to cross before we see the valley of minimum energy. In this case, it is much more difficult for the solution mechanism to find the global optimum. Processing is slow and the attentional parameters may become much more influential. That corresponds to the traits of controlled processing.

Levinson claims that GCIs are processed automatically. Though Levinson does not discuss the implementation of his default mechanism in connectionist networks, this can be done, for example, along the lines demonstrated in Blutner (2004). What is requires then is a system of constraints with strict domination. That is exactly what Levinson is proposing: Q >> M >> I. Hence Levinson's assumption about the ranking of the constraints fits together with his assumption about the automaticity of processing assuming the proposed picture of embodied cognition. However, I doubt the automaticity of processing in connection with GCIs and, consequently, I also doubt the strictness of domination between the pragmatic constraints Q, M and I. Hence, the position of RT – assuming controlled processing – is more appropriate when performance factors are under debate. The problem of finding the appropriate constraints is an open issue so far I can see. Neither RT nor Levinson's GCIs are fully adequate at the moment.

3.3 Neo-Gricean Approach

Neo-Griceans (Atlas & Levinson, 1981; Horn, 1984; Blutner, 1998; e.g. Atlas, 2005; Horn, 2005) are assuming two countervailing optimization principles: the Q-principle and the R-principle.[9] The first is oriented to the interests of the hearer and looks for optimal interpretations; the second is oriented to the interests of the speaker and looks for expressive optimization. Here is a standard presentation of the two principles (cf. Horn, 1984, 1989, 2004, 2005):

The Q-Principle (Hearer-based):
    Make your contribution sufficient!
    Say as much as you can! (modulo R)
    (Grice's first quantity maxim and the first two manner maxims)

---

[7] This corresponds to a strict domination of constraints. Strict domination means that a violation of an arbitrary number of lower ranked constraints can never overpower to violation of ONE higher ranked constraint.

[8] The standard stochastic algorithm is adiabatic annealing as used in the Boltzmann machine for connectionist networks (e.g. Rumelhart, Hinton, & McClelland, 1986). Biro (2006) has adapted the basic ideas for the use in OT.

[9] In OT, these 'principles' correspond to different directions of optimization where the *content* of the optimization procedure is expressed by particular OT constraints. This will be pointed out in more detail in the following section.

The R-Principle (Speaker-based):
    Make your contribution necessary!
    Say not more than you must! (modulo Q)
    (Grice's second quantity maxim, relation maxim and the second two manner maxims)

It is tempting to identify the Q-principle with Levinson's Q-heuristic and the R-principle with the I-heuristics. However, they are not identical though there is a correspondence between them. The difference has to do with the different status of *principles* in the global, neo-Gricean pragmatics on the one hand and *heuristics* (*defaults*) in Levinson's local pragmatics on the other hand. According to the neo-Gricean picture the principles constitute a kind of communication game – either between real speakers and hearers or between fictive speakers and hearers in the mind of a language user. In this game both principles are applied in a recursive way (corresponding to the modulo-clause in the formulation of the principles). In Levinson's theory, no such interaction between real or fictive speakers/hearers takes place. Instead, presumptive meanings are default interpretations and they are processed in a nearly automatic way. No 'mind reading' facilities are required here. The difference becomes quite clear when we give formalization in terms of unidirectional and bidirectional OT.

In section 2 I have introduced weak bidirection and it was illustrated how this solution concept explains Horn's division of pragmatic labour. If we assume that the optimization procedure is supplemented by a system of ranked constraints – in order to provide the content of the optimization – then Horn's R-principle/Q-principle is in exact correspondence to interpretive/expressive optimization. Further, the modulo-clause in the formulation of the Q-/R-principle is explicitly expressed by the recursive term in formalism as can be seen from the definition given in (3).

An important question concerns the *status* of the theory with regard to synchrony versus diachrony. Obviously, both RT and Levinson's theory of presumptive meanings take the synchronic view and both suggest a model of online language interpretation. Within the neo-Gricean camp, the situation is not so clearly decided. As mentioned in the introduction, prominent researchers take a normative stance. For instance, Horn (2005) claims the following:

> Grice's goal of developing an account of speaker meaning (of which implicature constitutes a proper subpart) is distinct from Relevance theorists' goal of developing a cognitive psychological model of utterance interpretation, which does not address the question of how and why the speaker, given what she wants to convey, utters what she utters. (194)

However, we agree with Carston (2005) that this statement is too strong as it stands. I think that the difference between the normative stance and the naturalistic stance should not be overestimated because in practice there is seldom a deep empirical conflict between the two stances (Spohn, 1993). Interestingly, the normative view comes very close to the diachronic perspective considering the Q and the I principle as diametrically opposed forces in language change. The diachronic perspective was clearly expressed in Horn (1984). Similarly, in the framework of OT pragmatics it is very natural to take weak bidirection as expressing a basic principle of natural language change. As a consequence, bidirectional optimization has nothing to do with online processes that run during normal language interpretation/production. Rather, the results of bidirectional optimization are routinized or fossilized – a phenomenon that takes place on an evolutionary time scale. Hendriks et al. (to appear) put this point as follows:

> On Blutner and Zeevat's evolutionary view of bidirectionality, form-meaning pairs that have been determined by bidirectional optimization constitute fixed relations to a

learner who sets out acquiring the language. No learner, indeed no user of the language, needs to perform a bidirectional computation for any form-meaning pair she encounters.

In contrast to this view there are representatives of OT pragmatics who suggest a procedural formulation of week bidirection and propose it as a realistic model of natural language interpretation and/or natural language production (e.g. Zeevat, 2000; Jäger, 2002; e.g. Beaver & Lee, 2004; Hendriks & Spenader, 2005/2006). This position is also taken in Hendriks et al. (to appear):

> However, we take the position that bidirectionality is not in the first place an evolutionary mechanism. Some form-meaning pairings have not been fossilized or automatized, but must be computed anew in a given situation. This view of bidirectionality raises the question of whether bidirectionality is a property of an individual's linguistic performance from the onset of language acquisition, or whether it is acquired or instantiated at some later time. We believe that the latter is the case. Whenever a bidirectional pair has to be computed online in a given situation, it is necessary for the hearer to realize which options were available to the speaker, and also to realize that the speaker's eventual choice is codetermined by the speaker's assumption that the hearer is able to share his perspective. It is to be expected that such online computation requires considerable cognitive resources.

In section 5 we will try to decide the issue by discussing recent empirical studies that relate to the two different positions.

3.4 Relations between the three frameworks

In Figure 2 I have illustrated two ways of describing Horn's division of pragmatic labour: (a) by assuming the two dominant linking constraints $F \rightarrow M$ and $*F \rightarrow *M$; (b) by assuming the markedness constraints F and M. In the first case unidirectional optimization gives already the intended solution (both in the interpretation perspective and the production perspective). In the second case it requires weak bidirectional optimization to get the intended solution, and this requires much more computational power than the unidirectional case. The significant computational complexity is compensated by the simplicity of the involved knowledge base since only markedness constraints are involved here. In the first case, contrastingly, the computational simplicity is compensated by the more complex system of linking constraints. Interestingly, there is a general way to transform the (b)-system into a (a)-system, and I have introduced the term *fossilization* for describing the relevant transfer. The corresponding mechanism of evolutionary learning will be discussed in the following section.

   Taken the example from Figure 2, Levinson's (2000) formulation corresponds to the (a)-system and Horn's (1984) formulation corresponds to the (b)-system. Correspondingly, we can see Levinson's system as a *fossilization* of Horn's neo-Gricean system. Hence, both systems are compatible if we take the different perspectives into account: the synchronic perspective for the (a)-system and the diachronic perspective for the (b) system.

   Though RT takes presumably another system of constraints than Levinson's, the RT system can likewise be seen as a fossilized processing architecture where no explicit mind reading capacities are required. Fossilization or routinization does not necessarily result in automatic processing (in contrast to Levinson's suggestion). I have discussed previously why the controlled character of processing assumed in RT is fully compatible with the neuronal embodiment of OT pragmatics.

   Given the existence of fossilization and the corresponding transfers, is can be concluded that the three discussed variations on Grice are much more closely related than the occasional

polemics let us expect. Despite of their similarities there are also important differences, mostly relating to different assumptions concerning the nature of natural language processing.

4 Fossilization: a bidirectional OT account

The idea of fossilization was introduced in Geis & Zwicky's (1971) paper about 'invited inferences' as a mechanism of conventionalization for implicatures. A closely related approach is Morgan's (1978) theory of *short-circuited implicatures* where some fundamentally pragmatic mechanism has become partially grammaticized. Leaning on this idea, Horn & Bayer (1984) propose an elegant account of so-called neg-raising, "the availability (with certain predicates) of lower clause understandings for higher clause negations" (p. 397). There is a principal difficulty for nonsyntactic treatments of these neg-raising interpretations. The difficulty has to do with the existence of lexical exceptions to neg-raising, i.e. we find pairs of virtual synonyms of which one member allows the lower clause understanding and the other blocks it.[10] Horn & Bayer (1984) argue that conversational implicatures may become conventionalized ("pragmatic conventions") and this conventionalization sanctions neg-raising. The short-circuiting of implicatures as a matter of convention has important empirical consequences, some of them we will discuss in the following section.[11]

In an early paper, Cole (1975) investigates similar phenomena in the lexical realm. Calling the conventionalization phenomenon "lexicalization of contextual meaning" he makes quite clear that the relevant conventions are built on the basis of particularized conversational implicatures (i.e. what Levinson (2000) calls *utterance token meanings*). Further, he proposes a diagnostics for discriminating between implicatures proper and their lexicalized counterpart. This may help to clarify the synchronic/diachronic status of conversational implicatures.

Traugott and her colleagues (e.g. Traugott, 1989; Traugott & Dasher, 2005) applied the idea of fossilization to explain language change. According to this model innovation may arise in the individual and spread or propagate through the community. In their *invited inferencing theory of semantic change*, Traugott and co. postulate a cycle starting with coded meaning, exploiting particularized conversational implicatures, transforming these implicatures into generalized conversational implicatures (= conventionalization), and finally resulting in new coded meanings (cf. Traugott & Dasher, 2005). Figure 3 shows a simplified picture of this model.

---

[10] One of Horn & Bayer's (1984) examples concerns opinion verbs. For instance, Hebrew *xogev* 'think' permits NR readings while *maamin* 'believe' does not. Interestingly, the opposite pattern obtains in Malagasy. In French *souhaiter* 'hope, wish' exhibits neg-raising, but its near-synonym *espdrer* does not – although it's Latin etymon *sperare* did. (cited after Horn & Bayer, 1984, p. 400).

[11] For example, we expect to find differences between speakers and between languages as to just which conventions of usage are operative. And exactly this happens as it is pointed out in Horn & Bayer (1984).
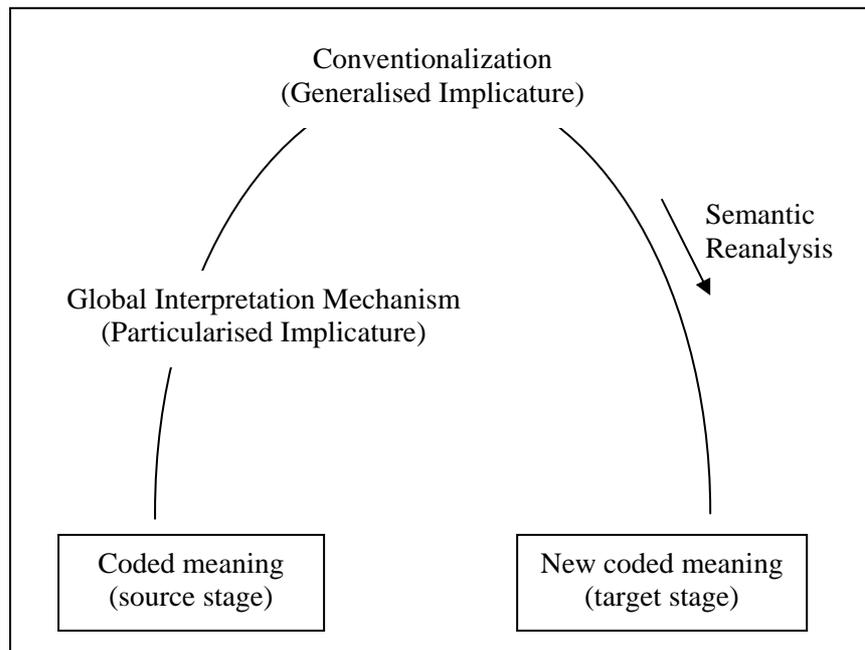
Figure 3. Simplified representation of the *invited inferencing theory of semantic change* (see Traugott & Dasher 2005: 38)

In the domain of syntax, Levinson (2000) und Mattausch (2004) used very the same idea for explaining the development of binding principles.

I will use the term *fossilization* here in a very broad sense that covers the whole spectrum of the mentioned phenomena. It stands for processes of *individual fossilization* that take place in individual language acquisition, i.e. on a time scale of seconds, hours and months. What's more it stands for social processes of *cultural fossilization* that take place in language change on a historical time scale of years up to centuries; the relevant mechanism is iterated learning / cultural evolution.

In OT pragmatics, fossilization relates to a transformation of knowledge systems. As we have seen (Figure 2) it is possible to describe the same solution space in two different ways. In the first case (Figure 2a) unidirectional optimization (either hearer or speaker perspective) is sufficient to calculate the solution pairs. It is plausible to assume that this kind of OT systems can be used to construct cognitively realistic models of online, incremental interpretation (see Blutner, 2006, 2007). The second case (Figure 2b) is using the recursion of weak bidirection (super-optimality) and has a completely different status. Because of its strictly non-local nature the proposed algorithms that calculate the super-optimal solutions do not even fit the simplest requirements of psychologically realistic models of online, incremental interpretation (Zeevat, 2000; Beaver & Lee, 2004).[12]

The proper understanding of weak bidirection and super-optimality relates best to an off-line mechanism that is based on bidirectional learning (Blutner, Borra, Lentz, Uijlings, & Zevenhuijzen, 2002; Benz, 2003; Van Rooy, 2004). In these approaches the solution concept of weak bidirection is considered as a principle describing the results of language change: super-optimal pairs emerge over time in language change. This relates to the view of Horn

---

[12] There are several arguments why bidirectional OT cannot yield an online mechanism of linguistic competence. Beaver & Lee (2004) argue that if more rounds of optimization are allowed, the bidirectional OT-model severely overgenerates in the sense that in later rounds peculiar new form-meaning pairs will emerge as winners. Before the Beaver & Lee paper, Zeevat (2000) argued against the symmetric view of OT pragmatics starting from the famous rat/rad problem and its pragmatic counterparts.

(1984) who considers the Q and the I principle as diametrically opposed forces in language change, and it conforms to the good old idea that synchronic structure is significantly informed by diachronic forces.

For the sake of illustration let's go back to our example illustrated in Figure 2. Let's assume a population of agents who realize speaker- and hearer strategies based exclusively on the markedness constraints F and M. In this population each content is expressed in the simplest way (f_1) and each expression is understood in the simplest way (m_1). Let's assume further that these agents communicate with each other. When agent x is in the speaker role and intends to express m_1, then expressive optimization yields f_1. Agent y is a hearer who receives f_1 and, according to interpretive optimization, he gets the interpretation m_1 – hence the hearer understands what the speaker intends: successful communication. Now assume the speaker wants to express m_2. With the same logic of optimization he will produce f_1 and the agent y interprets it as m_1. In this case, obviously, the communication is not successful. Now assume some kind of *adaptation* either by iterated learning or by some mutations of the ranked constraint system (including the linking constraints). According to this adaptation mechanism the expected 'utility' (how well they understand each other in the statistical mean) can improve in time. In that way a system that is evolving in time can be described including its special attractor dynamics. In each case there is a stabilizing final state that corresponds to the system of Figure 2a where the two Levinsonian (2000) constraints **I** (= [F→M]) and **M** (= [F→M]) outrank the rest of the constraints. It is precisely this system that reflects Horn's division of pragmatic labour. The only condition we have to assume is that the marked contents are less frequently expressed than the unmarked contents.[13]

Hence, the important insight is that a system that is exclusively based on markedness constraints such as in Figure 2b is evolutionary related to a system based on highly ranked linking constraints such as in Figure 2a. It is opportune to present some more details at this point. Our own simulation studies (Blutner et al., 2002) have provided the following results assuming the three different strategies illustrated in Figure 4. Here the *Horn-strategy* describes the famous pattern of iconicity (Horn's division of pragmatic labour). The *anti-Horn-strategy* describes a kind of anti-iconicity, and the Smolensky-strategy describes the presumed initial state of a learner where unmarked forms and unmarked meanings are preferred simultaneously.

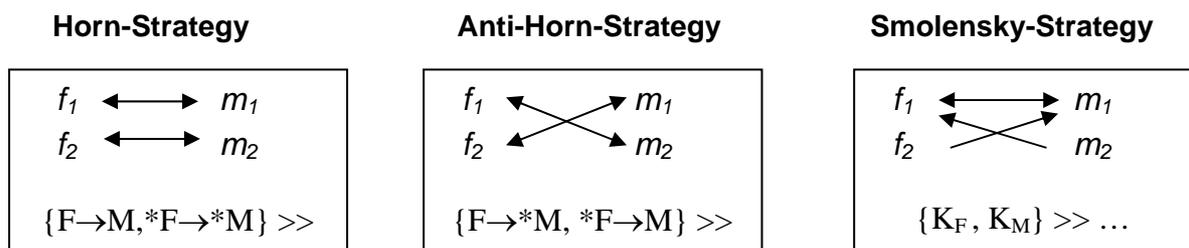| **Horn-Strategy**	**Anti-Horn-Strategy**	**Smolensky-Strategy**
$f_1 \longleftrightarrow m_1$ $f_2 \longleftrightarrow m_2$ {F→M,*F→*M} >>	$f_1 \quad m_1$ $f_2 \quad m_2$ (crossed arrows) {F→*M, *F→M} >>	$f_1 \longleftrightarrow m_1$ $f_2 \quad m_2$ (crossed arrows) {K_F, K_M} >> …

Figure 4: Three different strategies – based on the indicated three different rankings of the constraints

- Horn- und anti-Horn-strategies are the only evolutionary stable strategies.
- If the initial state represents a uniform Smolensky-population, then the systems develops into

---

- – a pure Horn-population, assumed the frequency of the realization of m_1 is higher than that of m_2: $P(m_1) > P(m_2)$
  - – a pure anti-Horn-population, assumed the frequency of the realization of m_2 is higher than that of m_1: $P(m_2) > P(m_1)$
- ▪ The corresponding proposition is true if the initial state represents a mixed population

Hence, the probabilities for the situations that are described, i.e. $P(m_1)$, $P(m_2)$, are decisive for the result. The classical pattern of iconicity is predicted only in cases where the unmarked situation has the highest probability. McCawley (1978) has listed numerous cases of constructional iconicity in the lexicon. Krifka (2007) has observed that the phenomenon is the decisive factor in determining the precise/vague interpretation of measure expressions.

Interestingly, there are also examples of anti-iconicity. They are found in connection with semantic broadening where the initial meaning is described as that of an ideal shape, figure or state. A good example can be found in Dutch, where besides the preposition *om* (= Engl. *round;* German *um*) the expressions *rond* and *rondom* are in use. The expression *rond* is a word borrowed from French. It refers to the ideal shape of a circle. Starting with its appearance it comes in competition with the original (and *unmarked*) expression *om*. The results is a division of labour as demonstrated in the following examples (cf. Zwarts, 2003, 2006):

(4) a. Ze zaten rond (?om) de televisie
        They sat round the television
    b. Een man stak zijn hoofd om (?rond, ?rondom) de deur
        A man put his head round the door
    c. De auto reed om (?rond, ?rondom) het obstakel heen
        The drove round the obstacle
    d. het gebied rondom (?om) het stadje
        the area round the little town

According to the principle of iconicity we would expect that the unmarked form (*om*) is paired with the ideal of the circle shape and the marked form (*rond*) with the detour interpretation.[14] However, the opposite is true. I think there is a simple explanation for this fact: ideal shapes/situations are much less frequent then non-ideal situations; hence, since $P(m_1) < P(m_2)$, the present evolutionary approach predicts anti-iconicity.

I think these examples and many other examples in the area of lexical pragmatics (e.g. Blutner, 1998; Wilson, 2003) strongly suggest the reality of fossilization. Accepting that both possibilities are real to some extend – the online calculation of implicatures and the access of their fossilized counterparts, the question concerns the distribution of labor between online processes (calculating optimal outcomes) and offline processing (fossilization processes). We can ask this question for standard scalar implicatures, as well as for other types of pragmatic inferences. In the next section I will review some experiments that are claimed to decide the issue. These experiments are closely related to the issue of asymmetries between comprehension and production processes.

---

[14] The assumption that the ideal path description (circle) is realizing the unmarked interpretation and the detour interpretation is realizing the marked interpretation is justified by independent thoughts about the preference of the logically strongest interpretation (e.g. Dalrymple, Kanazawa, Kim, Mchombo, & Peters, 1998).

5 Asymmetries between natural language comprehension and production

It's a common observation that we often are not able to produce what we can understand. The opposite situation, where we are able to produce a certain expression but unable to understand, it is observed much less often. The phenomenon of aphasia gives a feasible illustration of the existence of both kinds of asymmetries (e.g. Jakobson, 1941/1968). Likewise, in the domain of language acquisition both sides of the phenomenon can be detected. It is well known that children's ability in production lags dramatically behind their ability in comprehension (e.g. Benedict, 1979; Clark, 1993). It was only recently that attention was devoted to the opposite case where children's comprehension performance lags years behind their ability of production (cf. Hendriks & Spenader, 2005/2006) .

There are three different ways to deal with these observations. The first approach is to assume dissociation between a comprehension grammar and a production grammar. Unfortunately, this account requires some ad hoc stipulations which conflict with general assumptions of parsimony.

The second approach is to assume different processing restrictions for production and comprehension. Joshi (1987) was possibly the first who discussed the asymmetry issue from the viewpoint of artificial intelligence:

> Comprehension and generation, when viewed as functions mapping from utterances to meanings and intentions and vice versa, can certainly be regarded as inverses of each other. However, these functions are enormously complex and therefore, although at the global level they are inverses of each other, the inverse transformation (i.e., computation of one function from the other) is not likely to be so direct. So, in this sense, there may be an asymmetry between comprehension and generation even at the theoretical level. (Joshi 1987, p. 183)

Joshi further suggests (p. 184) that the human generation mechanism involves some monitoring of the output, presumably by the comprehension mechanism. A corresponding monitoring (by generation) is not assumed for the human comprehension mechanism.

The third way of dealing with the asymmetry follows from an optimization approach. This was first demonstrated by Smolensky (1996). As we have seen in the previous sections, natural language production in OT goes from a given interpretation to an optimal expression and natural language comprehension goes from a given expression to an optimal interpretation. It is these different directions of optimization which impose different boundary conditions on the process of optimization. As a result, the same system of constraints and the same constraint hierarchy can account for the observed asymmetry, without taking recourse to multiple grammars or different processing restrictions for production and comprehension.

In this section I will discuss asymmetries between comprehension and production in the context of recent experimentation. The natural language expressions investigated are pronouns, reflexives, referential and quantifying expressions – the latter in connection with scalar implicatures. The fundamental questions asked are twofold:

(i) How to explain the observed differences between comprehension and production in a certain stage of development?
(ii) What is the mechanism that handles how to overcome the gap between comprehension and production during natural language acquisition?

OT has a very simple answer to the question (i). In order to account for the usual observation that comprehension can be perfect while production is not, Smolensky (1996) assumes two kinds of constraints: (a) markedness constraints for forms and (b) linking (faithfulness) constraints – linking forms and meanings in an adequate way. Further, he assumes that the

markedness constraints initially dominate the linking constraints. It is exactly under these conditions that we get the expected pattern.

For sake of illustration, let us go back to the earlier example with two forms and two meanings. We assume the markedness constraint for forms F and the two linking constraints F→M and *F→*M (see table 1). If {F} >> {F→M, *F→*M} then comprehension is always correct (interpreting f_1 as m_1 and f_2 as m_2). However, the production perspective sometimes gives the wrong result. This is because of the dominance of the markedness constraint F, and it gives the result that all meanings m_i (i = 1,2) are expressed by the simpler form f_1.

Interestingly, the opposite pattern of delayed comprehension is also possible. In this case we have to assume an incomplete system of linking constraints that outranks the system of markedness constraints. A very simple example is {F→M} >> {F}. In this case m_1 produces f_1 and m_2 produces f_2. However, while f_1 is always interpreted correctly as m_1 the form f_2 comes out as ambiguous. It can be interpreted both as m_1 and m_2, and this constitutes a case of delayed comprehension.

The research question (ii) is much more difficult to answer. The difficulty arises from the fact that there is not only one potential mechanism to overcome the gap between comprehension and production. There are at least two such mechanisms, and I will consider them in correspondence with the two ways of viewing bidirection discussed earlier. The first mechanism is based on an OT learning mechanism that re-ranks the involved constraints. That's exactly Smolensky's view as taken in Smolensky (1996). The second mechanism is a mechanism of maturation resulting in a processing system that integrates the comprehension and the production perspective. The resulting integrated system can be either the symmetric system of bidirectional OT or an asymmetric version such as proposed by Joshi and worked out by Zeevat (2000).

In a slightly different formulation, the first mechanism is realizing the diachronic view of bidirection where bidirectional optimization takes place offline (during language acquisition) and leads to some kind of fossilizing optimal form-meaning pairs. In contrast, the second mechanism presumes bidirectional optimization as a psychologically realistic online mechanism. According to this online/synchronic view, speakers (hearers) optimize bidirectionally and take into account hearers (speakers) when selecting (interpreting) a referring expression. In the following I will consider some experimental investigations that shed a light on the empirical adequacy of these two positions.

5.1 The Pronoun Interpretation Problem

In a recent research article Hendriks & Spenader (2005/2006) give a new interpretation of children's delay of the comprehension of pronouns (see also Hendriks, Rijn, & Valkenier, 2007). I discuss the validity of their interpretation and present an alternative account in terms of iterated learning.

A series of experiments has shown that children make errors in interpreting pronouns as late as age 6;6, yet correctly comprehend reflexives from the age of 3;0 (e.g. Chien & Wexler, 1990; McKee, 1992; Koster, 1993; Spenader, Smits, & Hendriks, 2007). For example, children were confronted with a context where two boys, Bert and Paul, are introduced, and the following sentences were given:

(5) a. Bert is washing himself
    b. Bert is washing him

Sentences like (5a) are correctly understood from a young age (95% of the time according to some studies). However, children misinterpret the pronoun in (5b) as coreferring with the subject about half the time. Hence, it seems that children did not yet realize that the

coreferring reading of (5b) must be blocked given the existence of the sentence (5a) which clearly has the coreferring reading.

Contrasting with the comprehension data, language production experiments consistently have shown that children do not have problems in producing reflexives or pronouns correctly. For example, Bloom et al. (1994) demonstrated that even in the youngest age groups investigated (ranging from 2;3 to 3;10) the children consistently used the pronoun to express a disjoint meaning, while they used the reflexive to express a coreferential interpretation. It can be concluded from the production data that children have competence of binding principles. Why they don't use this knowledge in comprehension?

I cannot go into all the different theoretical proposals concerning the pronoun interpretation problem. Instead, I will be mainly concentrated on the possibilities opened by OT pragmatics. Recently, several authors have agued that the observed delay in comprehension can be explained by assuming that children are only able to consider their own perspective, whereas adult hearers are able to simultaneously take into account the perspective of the speaker (deHoop & Kramer, 2005/2006; Hendriks & Spenader, 2005/2006; Hendriks, Rijn et al., 2007).

As explained at the beginning of this section it is possible to account for the delay of comprehension by assuming an incomplete system of linking constraints that outranks the system of markedness constraints for forms, for instance the system $\{F{\rightarrow}M\} \gg \{F\}$. In the concrete case of pronoun/reflexive interpretation f_1 stands for the reflexive, f_2 for the pronoun, m_1 for the coreferential interpretation and m_2 for the disjoint interpretation. The markedness constraint F prefers he reflexive over the pronoun and can be read as "referential ecomomy" (see Burzio, 1998). The linking constrain $F{\rightarrow}M$ excludes the reflexive from the disjoint interpretation – that's just the binding principle A (a reflexive must be bound locally) expressed as a violable constraint. Figure 5 shows the preferences between the four possible form meaning pairs arising from the system.
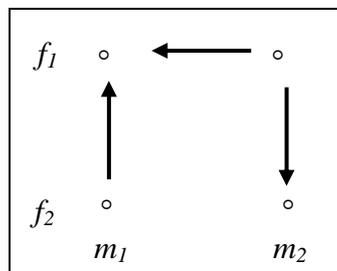

Figure 5: Diagram illustrating the pronoun interpretation problem. It shows the preferences between the four form-interpretation pairs based on the system $\{F{\rightarrow}M\} \gg \{F\}$ of ranked constraints

Assume now that children begin with unidirectional optimization. In the case of production everything goes right: the meanings m_1 and m_2 are expressed by f_1 and f_2, respectively. However, in case of comprehension the form f_2 (the pronoun) exhibits an ambiguity: both he interpretation m_1 and m_2 are optimal taken the interpretive perspective for optimization. And that's exactly the expression of the pronoun interpretation problem.

Optimizing bidirectionally inherently involves reasoning about alternatives not present in the current situation. In the present case a child who is hearing f_2 (a pronoun) must reason what other non-expressed forms the speaker could have used. It can realize then that a coreferential meaning m_1 is better expressed with f_1 (a reflexive). Then, by a process of elimination, the child must realize the pronoun should be interpreted as disjoint meaning m_2

and this resolves the ambiguity. Since the ability to optimize bidirectionally may be a skill acquired relatively late, this idea gives a plausible explanation of the lag in acquisition.

Summarizing, the online processing account of Hendriks & Spenader (2005/2006) provides a new way to explain children's delay of the comprehension of pronouns. What's essential for this solution is the hypothesis that the hearer has to take a potential speaker into account. Thus, the authors are able to derive principle B effects (pronouns are free) from principle A alone, through bidirectional optimization. The approach nicely combines a pragmatic explanation with a processing account (lack of processing resourses). Besides the stipulation of the constraints and their ranking no other stipulations are required.

However, there are also some arguments that challenge the discussed view. First at all there is the question of constraint grounding. Other systems of constraints are conceivable and successfully used in the literature (see, e.g., Levinson, 2000; Mattausch, 2004). Further, there is no answer on the question why the particular ranking {F→M} >> {F} is assumed. Another problem has to do with children's abilities for mind reading ("theory of mind") that is explicitly assumed in Hendriks' and Spenader's approach. The assumption of mind reading as a prerequisite for making the transition to bidirectional reasoning has the consequence that there should be strict correlations between the behavior in standard tests of theories of mind (see Perner, Leekam, & Wimmer, 1987) and the behavior in tasks involving bidirectionality (such as pronoun interpretation). Unfortunately, such strict correlations never were found (Flobbe, Verbrugge, Hendriks, & Krämer, 2007). Further, mind reading requires awareness of other conversation participant's choices. Hence, it is based on controlled rather than automatic processing. However, pronoun processing appears to be automatic rather than controlled. There is no explicit hint for mind reading capacities in such tasks.

In the following subsection I will propose an alternative account that can describe the same kind of data and in addition has some conceptual advantages.

5.2 Pronoun interpretation and related task: individual fossilization

In section 4 I described an approach to fossilization and I made a distinction between individual and cultural fossilization. Cultural fossilization was successfully used by Mattausch and Jäger (Jäger, 2004; Mattausch, 2004). I will consider now individual fossilization in connection with the pronoun interpretation problem.

In the informal description given here the focus is on pointing out the differences to the processing account provided by Hendriks & co. Let's start with Hendrik's initial system {F→M} >> {F}. In order to apply OT learning theory we assume that a complete system of constrains is present in a background of equally ranked constraints. The following system which is functionally equivalent with the system described before is used: {F→M} >> {F} >> {F→*M, *F→M *F→*M}. The learning rule then says: promote constraints that favour wanted behaviour over unwanted, demote constraints that favour unwanted behaviour over wanted. If a competent adult acts as speaker and the child as hearer, then this learning rules lead to the promotion of *F→*M (principle B). Figure 6 illustrates the transfer between the two systems.[15]

---

[15] Alternatively, we could start with the system {*F, *M} >> {F→M, F→*M, *F→M, *F→*M}. The two dominating constraint *F and *M express that f_2 (pronoun) is the preferred form and that m_2 (disjoint interpretation) is the preferred interpretation. The linking constraints cancel each other. Then it can be shown that iterated learning leads to different stages of development. First the principle A is evolving if the plausible stipulation is made that $P(m_2) > P(m_1)$. Hence we have a motivation why the system of preferences as given on the left hand site of Figure 6 appears – it reflects delayed comprehension – instead of a system exhibiting delayed production. Only later the principle B becomes dominant, giving the preferences shown on the right hand site of Figure 6.
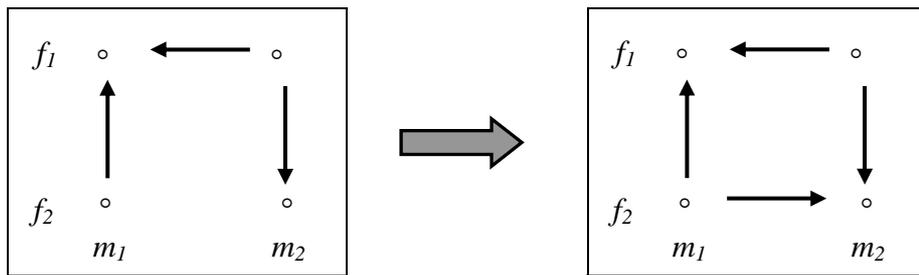
Figure 6: Transformation between two systems of ranked constraints provided by individual fossilization

It is obvious that this transfer is dependent of parameters of use. Hence, we expect frequency effects when the fossilization mechanism is at work. Further, we would expect no significant differences between the comprehension of pronouns and the comprehension of reflexives. The reason is that their processing loads are not significantly different. Hendriks' online view of processing (involving bidirection) conflicts with both hypotheses. It suggests a domain-independent transition from the unidirectional to the bidirectional case. Consequently, we shouldn't expect significant effects of use (frequency effects). Further, for adult subjects we should expect significant differences in processing between pronouns and reflexives, since the pronoun requires bidirectional processing but the reflexive does not.

I think both hypotheses supporting the fossilization view can be confirmed. Though there is no direct verification of the second hypothesis at the moment, I think in the light of the eye tracking investigations of Karabanov, Bosch, & König (to appear) it is not probable that the comprehension of pronouns takes significantly more time than the comprehension of reflexives assuming comparable conditions. For the first hypothesis, it's important to see that there are some other domains which realize the same structural relations as exhibited in the case of pronoun interpretation exhibited in Figure 5. Consider first the domain of natural language quantifiers and consider dual quantifiers such as *some(A)* and *all(A)*, where A stand for a certain restrictive term. Logically, *all(A)(B)* has the set inclusion interpretation stating $A \subseteq B$, and *some(A)(B)* has an interpretation expressing nonempty intersection $A \cap B \neq \varnothing$. Of course, this interpretation does not exclude the set inclusion reading. It's the scalar implicature that excludes this interpretation – leading to a *some_but_not_all* interpretation. The ordering of all form interpretation pairs given in Figure 5 can be applied to the quantifier case when we assume that f_1 stands for *all*, f_2 for *some*, m_1 for the set inclusion interpretation, and m_2 for the *some_but_not_all* reading. The markedness constraint F now prefers *all* over *some*. We can see that as a realization of the strongest meaning hypothesis (Dalrymple et al., 1998). Further, the dominating constraint F→M expresses the meaning postulate for *all*, and the potential constraint *F→*M expresses the scalar implicature for *some*.

The first systematic investigation of the acquisition of scalar implicature can be attributed to Noveck (2001). From his experiments it can be concluded that young children initially treat a relatively weak term logically before becoming aware of its pragmatic potential, and that, in this respect, "children are more logical than adults" (Noveck, 2001: 165). Concluding, we can speak of delayed comprehension of the pragmatic potential of the weak quantifier.

Another domain where we find similar effects is the interpretation indefinite expressions. In several languages it has been observed that indefinite noun phrases such as *a boy* take on different interpretations depending on whether they appear in a scrambled or unscrambled word ordering (e.g. De Hoop & Krämer, 2005; Unsworth, 2005). Adults interpret unscrambled indefinites (f_1) as 'non-specific' (m_1) whereas they interpret scrambled indefinites (f_2) as 'specific' (m_2). Again we find a delayed comprehension effect: children

interpret scrambled indefinites in both ways. Only later they realize that the 'specific' interpretation is the proper one.

In a recent article Hendriks et al. (2007) discuss the results of diverse experiments in different domains and conclude that children seem to differ in the ages at which they provide adult-like responses for particular linguistic forms.

> Whereas from the age of 6 or 7 on children start to interpret pronouns correctly, children until roughly 11 years old select a non-adult meaning for indefinite objects (Unsworth, 2005), and many 10- and 11-year-olds do not draw a scalar implicature where most adults would (Noveck, 2001). This suggests that bi-directional optimization is not a general strategy that has to be learned by children in one step, but rather that the possibility of bi-directional optimization is dependent on the frequency of use of the relevant production rules. (p. 1893)

Hence, the first hypothesis suggested above – predicting a domain-independent transition from the unidirectional to the bidirectional view – seems to be falsified. And this might be a powerful argument supporting the fossilization view.

Thought the domain independence of he transfer from unidirectional to bidirectional processing is a natural consequence of the online processing view, it is not a necessary consequence. Hendriks et al. (2007) provide an improvement of their online processing view in order to describe the empirically found domain dependency. This improvement is formulated in terms of the ACT-R model (Anderson & Lebiere, 1998; Anderson et al., 2004).

ACT-R understands itself as an integrated theory of the mind. Different from Smolensky's (Smolensky & Legendre, 2006) theory of *harmonic mind* which sees the symbolic part (i.e., OT) as a high-level description of the neural realm, ACT-R is a *hybrid* theory that relates different symbolic modules with certain subsymbolic processes. These subsymbolic processes serve to guide the selection of rules to fire as well as the internal operations of modules and much of learning.

Hendriks et al. (2007) model unidirectional and bidirectional OT in terms of the ACT-R model. In this model bidirectional optimization is described as the serial application of two unidirectional processes of optimization. A crucial property of ACT-R is the assumption that actions take time to perform and that performance is limited by the serial processing bottleneck. Since bidirectional optimization needs much more processing resources than unidirectional optimization does, a process of production compilation[16] comes in increasing the processing efficiency. The result of product compilation conforms to an instance based kind of automatization (Logan, 1988). I think what is described here comes very close to the idea of fossilization. Whereas fossilization leads to the introduction of new linking constraints product compilation leads to the generation of new productions who describe the results of certain bidirectional actions.

5.3 Choosing the right referring expression

The standard case of production/comprehension asymmetries is delayed production. Comprehension can be perfect while production is not. A good example is given by production and understanding of R-expressions and pronouns as illustrated in (6).

---

[16] In production compilation, two existing production rules are integrated into one new production rule. Production compilation occurs when two existing production rules are repetitively executed in sequence.

(6) Discourse context: A woman is waiting at the corner. Her girl is eating an ice cream cone.
    a. She wears a red shirt.
    b. The woman wears a red shirt.

The interpretation of the pronoun in (6a) clearly refers to the discourse topic (*the girl*). If we want to express the alternative meaning as in (6b) we cannot use the pronoun. Interestingly, young children very often produce such subject pronouns when intending to refer to non-topics. Karmiloff-Smith (1985) found this pattern of production in children until the age of 6.

    I have already mentioned that the phenomenon can be modeled by assuming markedness conventions that initially dominate linking constraints. Figure 7 shows the corresponding diagram. Hereby, f_1 stands for the pronoun and f_2 for an R-expression. Further, m_1 is the interpretation referring to the topicalized discourse referent while m_2 refers to the non-topicalized one. F can be seen as referential economy (preferring pronouns to R-expressions) and F$\rightarrow$M expresses the preference for pronouns to be interpreted as the topic of the discourse.
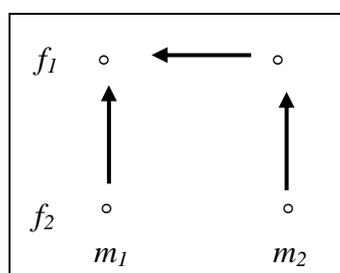

Figure 6: Diagram illustrating the pronoun production problem. It shows the preferences between the four form-interpretation pairs based on the system {F} >> {F$\rightarrow$M} of ranked constraints

Using unidirectional optimization, the diagram describes the OT system of an agent who can properly understand pronouns and R-expressions but who overuse pronouns when intending to refer to non-topics.

    The two considered models now make different assumptions for describing the transfer from the child system to the adult system. The online processing model handles the production problem by assuming that the producer takes the hearer into account and begins to reason bidirectionally at some point of development. In contrast, the fossilization view says that unidirectional optimization is sufficient if it is assumed that the relevant information has been fossilized at some part of the human development.

    In a recent research article, Hendriks, Englert, & Wubs (2007) argue that the investigation of elderly adults could decide between the two models. Elderly adults possess the required pragmatic and grammatical knowledge to select and interpret referring. However, their linguistic performance can be defective, due to a decreasing working memory capacity. And indeed, the authors found that elderly adults produce non-recoverable pronouns significantly more often than young adults when referring to the old topic in the presence of a new topic. With respect to the comprehension task, no significant differences were found between elderly and young adults.

    Obviously, this experimental outcome is a great problem for the fossilization view, since a stipulation of a mechanism of 'de-fossilization' does not make any sense in the present context. Hence, the assumption that the speaker takes the hearer into account is well

motivated for such examples. Zeevat (2000) has argued for this kind of active, creative processes.

However, there is also a problem for the bidirectional processing view. It says that both the speaker takes the hearer into account and, vice versa, the hearer takes the speaker into account. If that is right, then the same argumentation that is given in the paper by Hendriks, Englert, & Wubs (2007) can be applied for the delayed comprehension experiments discussed in the previous subsections. Thought I don't know of any experiments with elderly people concerning the delayed comprehension task, I bet more than my finger that the behavior of elderly people does not go down to that of young children in the relevant respects. Hence what we can conclude from these experiments is an asymmetry of processing: the speaker takes the hearer into account but not necessarily vice versa. This is actually Zeevat's (2000) view of making a distinction between the active and creative process of production and the rather passive process of interpretation.[17] The idea of fossilization is needed in order to account for the delayed comprehension data.

6 Conclusions

The aim of this article was to close the gap between experimental pragmatics and neo-Gricean theories of pragmatics as formulated in OT pragmatics. I have argued that OT pragmatics has the potential to account both for the synchronic and the diachronic perspective in pragmatics. I further have pointed out that the concept of fossilization can help to understand the idea of naturalization and (cultural) embodiment in the context of natural language interpretation.

In discussing modern pragmatic theories I have shown that RT and Levinson's theory of presumptive meanings account for the resolution of the conflict between effort minimization and effect maximization in different ways. In a certain sense, the crux of both approaches can be translated in OT pragmatics by making use of particular linking constraints. This translation makes the advantage of both approaches visible: both conform to the incremental, online character of natural language interpretation.[18]

Regarding the theoretical status of fossilization, there are important open questions. In a by now classical paper Cole (1975) considered the following example of a true conversational implicature, where a girl called Pamela upon being asked (7) might reply (8):

(7) How are you doing in your new position at San Andreas Fault University?

(8) Well, I haven't been fired yet.

Although the logical content of (8) is roughly that of the proposition that Pamela has not yet lost her job, more than that is implicated, namely that Pamela is not doing well. In this example, the implicature is really novel. There is no construction involved whose frequent use

---

[17] "The situation can be fruitfully compared to the habit of hiding easter eggs for one's children. The parents engaged in hiding the eggs balance the amount of effort with the desired amount of difficulty in finding the egg. (They also picture the child looking for it and try to keep it possible for the child of finding the egg, without spoiling the fun.) For the child it is another matter. They just have to throw in the effort required for finding the eggs. Not more of course, but definitely not less. It is not a complicated balancing act." (Zeevat 2000: 245)

[18] In discussing processing characteristics, incrementality and automaticity of processing have to be discriminated. Whereas automaticity of processing implicates the incremental character of processing the opposite is not true: incrementality does not implicate automatic processing. RT explains the incremental character of processing and has good reasons for assuming controlled processing in order to account for the processing of conversational implicatures. That's different from Levinson's account which assumes automatic processing for generalized conversational implicatures. It seems that RT is better justified on empirical grounds (cf. Noveck & Sperber, 2005).

could lead to the fossilization phenomenon (Cole's term is 'lexicalization'). Hence, this implicature is different from many other cases where a certain amount of fossilization is plausible. The important question is how to discriminate between offline implicatures that are not fossilized and their fossilized counterparts. Where is the boundary between aspects of interpretations that are truly conversational and aspects which have become lexically (or syntactically) encoded? We think the former aspect of interpretation can require some real mind reading capacities, requires conscious reflections and proceeds offline. So far we can see none of the discussed pragmatic theories has an interesting answer for this long-standing and intriguing question.

In the last part of the paper I have discussed recent work about the phenomenon of delayed comprehension and delayed production. This is a phenomenon which was not discussed within experimental pragmatics, though the importance of the problem was clearly recognized within OT pragmatics. I have discussed two models which conceptualized bidirection in different ways: the online processing model and the fossilization account. I have argued that neither of these extreme views gives a complete fit to the empirical data when taken per se. While it is obvious that fossilization phenomena are real to some extent it can be argued that a restricted online version of bidirection is correct: speakers optimize bidirectionally and take the hearer into account when calculating the optimal expression; in contrast, hearers normally do not take the speaker into account when calculating the optimal interpretation.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060.

Anderson, J. R., & Lebiere, C. (1998). *The Atomic Components of Thought*. Mahwah, NJ.: Erlbaum.

Anderson, M. L. (2003). Embodied Cognition: A field guide. *Artificial Intelligence, 149*, 91–130.

Atlas, J. D. (2005). *Logic, Meaning, and Conversation: Semantical Underdeterminacy, Implicature, and Their Interface*. Oxford: Oxford University Press.

Atlas, J. D., & Levinson, S. C. (1981). It-clefts, informativeness and logical form. In P. Cole (Ed.), *Radical Pragmatics* (pp. 1-61). New York: Academic Press.

Axelrod, R. (1984). *The evolution of co-operation*. London: Penguin.

Beaver, D., & Lee, H. (2004). Input-output mismatches in OT. In Palgrave/Macmillan (Ed.), *Optimality Theory and Pragmatics*. Houndmills, Basingstoke, Hampshire.

Benedict, H. (1979). Early Lexical Development: Comprehension and Production. *Journal of Child Language, 6*, 183-200.

Benz, A. (2003). Partial Blocking, associative learning, and the principle of weak optimality. In J. Spenader & A. Eriksson & Ö. Dahl (Eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory* (pp. 150-159). Stockholm.

Biro, T. (2006). *Finding the right words: Implementing Optimality Theory with simulated annealing.* University of Groningen, Groningen.

Bloom, P., Barss, A., Nicol, J., & Conway, L. (1994). Children's knowledge of binding and coreference: Evidence from spontaneous speech. *Language, 70*(1), 53-71.

Blutner, R. (1998). Lexical pragmatics. *Journal of Semantics, 15*, 115-162.

Blutner, R. (2000). Some aspects of optimality in natural language interpretation. *Journal of Semantics, 17*, 189-216.

Blutner, R. (2004). Nonmonotonic inferences and neural networks. *Synthese (Special issue Knowledge, Rationality and Action), 142*(2), 143-174.

Blutner, R. (2006). Embedded implicatures and optimality theoretic pragmatics. In Torgim Solstad & A. Grønn & D. Haug (Eds.), *A Festschrift for Kjell Johan Sæbø: in partial fulfilment of the requirements for the celebration of his 50th birthday*. Oslo.

Blutner, R. (2007). Optimality Theoretic Pragmatics and the Explicature/Implicature Distinction. In N. Burton-Roberts (Ed.), *Advances in Pragmatics* (pp. 67-89). Houndmills, Basingstoke, Hampshire: Palgrave/MacMillan.

Blutner, R., Borra, E., Lentz, T., Uijlings, A., & Zevenhuijzen, R. (2002). Signalling games: Hoe evolutie optimale strategieen selecteert, *Handelingen van de 24ste Nederlands-Vlaamse Filosofiedag*. Amsterdam: Universiteit van Amsterdam.

Blutner, R., de Hoop, H., & Hendriks, P. (2005). *Optimal Communication*: CSLI.

Blutner, R., & Zeevat, H. (Eds.). (2004). *Optimality Theory and Pragmatics*. Houndmills, Basingstoke, Hampshire: Palgrave/Macmillan.

Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry, 32*, 45-86.

Burzio, L. (1998). Anaphora and soft constraints. In P. Barbosa & D. Fox & P. Hagstrom & M. McGinnis & D. Pesetsky (Eds.), *Is the best good enough?* Cambridge, Mass.: The MIT Press.

Carston, R. (2002). *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Carston, R. (2005). Relevance Theory, Grice and the neo-Griceans: a response to Laurence Horn's 'Current issues in neo-Gricean pragmatics'. *Intercultural Pragmatics, 2*(3), 303-319.

Chien, Y.-C., & Wexler, K. (1990). Children's knowledge of locality conditions on binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition, 13*, 225-295.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and Beyond* (pp. 39-103). Oxford: Oxford University Press.

Clark, E. V. (1993). *The Lexicon in Acquisition*. Cambridge: Cambridge University Press.

Cole, P. (1975). The synchronic and diachronic status of conversational implicature. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics, Volume 3: Speech Acts* (pp. 257-288). San Diego, Cal.: Academic Press, Inc.

Dalrymple, M., Kanazawa, M., Kim, Y., Mchombo, S., & Peters, S. (1998). Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy, 21*, 159-210.

de Hoop, H., & de Swart, H. (2000). Temporal adjunct clauses in optimality theory. *Rivista di Linguistica, 12*(1), 107-127.

De Hoop, H., & Krämer, I. (2005). Children's Optimal Interpretations of Indefinite Subjects and Objects. *Language Acquisition, 13*, 103–123.

deHoop, H., & Kramer, I. (2005/2006). Children's Optimal Interpretations of Indefinite Subjects and Objects. *Language Acquisition, 13*(2), 103-123.

Dekker, P., & van Rooy, R. (2000). Bi-directional optimality theory: An application of game theory. *Journal of Semantics, 17*, 217-242.

Flobbe, L., Verbrugge, R., Hendriks, P., & Krämer, I. (2007). *Children's application of Theory of Mind in reasoning and language*.Unpublished manuscript, University of Groningen & Radboud University Nijmegen.

Fodor, J. (1975). *The Language of Thought*. New York: Thomas Crowell.

Gazdar, G. (1979). *Pragmatics*. New York: Academic Press.

Geis, M., & Zwicky, A. (1971). On invited inference. *Linguistic Inquiry, 2*, 561-579.

Harnad, S. (1990). The symbolic grounding problem. *Physica D, 42*, 335-346.

Hendriks, P., & de Hoop, H. (2001). Optimality theoretic semantics. *Linguistics and Philosophy, 24*, 1-32.

Hendriks, P., Englert, C., & Wubs, E. (2007). *Age differences in adults' use of referring expressions.*Unpublished manuscript, University of Groningen.

Hendriks, P., Hoop, H. d., Krämer, I., Swart, H. d., & Zwarts, J. (to appear). *Conflicts in Interpretation*.

Hendriks, P., Rijn, H. v., & Valkenier, B. (2007). Learning to reason about speakers' alternatives in sentence comprehension: A computational account. *Lingua, 117*, 1879–1896.

Hendriks, P., & Spenader, J. (2005/2006). When production precedes comprehension: An optimization approach to the acquisition of pronouns. *Language Acquisition, 13*(4), 319-348.

Hobbs, J., & Martin, P. (1987). Local Pragmatics, *Proceedings, International Joint Conference on Artificial Intelligence* (pp. 520-523). Milan.

Horn, L. (1984). Towards a new taxonomy of pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context: Linguistic applications* (pp. 11-42). Washington: Georgetown University Press.

Horn, L. (1989). *A natural history of negation*. Chicago: Chicago University Press.

Horn, L. (2004). Implicature. In L. Horn & G. Ward (Eds.), *Handbook of Pragmatics*. Oxford: Blackwell.

Horn, L. (2005). Current issues in neo-Gricean pragmatics. *Intercultural Pragmatics, 2*(2), 191-204.

Horn, L., & Bayer, S. (1984). Short-circuited implicature: A negative contribution. *Linguistics and Philosophy, 7*, 397-414.

Jäger, G. (2002). Some notes on the formal properties of bidirectional optimality theory. *Journal of Logic, Language and Information, 11*, 427-451.

Jäger, G. (2004). Learning constraint sub-hierarchies. The bidirectional gradual learning Algorithm. In R. Blutner & H. Zeevat (Eds.), *Pragmatics and Optimality Theory*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.

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

Jaszczolt, K. M. (to appear). Semantics and Pragmatics: The Boundary Issue. In C. Maienborn & K. v. Heusinger & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning*. Berlin: Mouton de Gruyter.

Joshi, A. K. (1987). Generation - a new frontier of natural language processing? *Theoretical Issues in Natural Language Processing, 3*, 181-184.

Karabanov, A., Bosch, P., & König, P. (to appear). Eye Tracking as a tool for investigating the comprehension of referential expressions. In S. Featherston & W. Sternefeld (Eds.), *Roots: Linguistics in search of its evidential base*. Berlin: De Gruyter.

Karmiloff-Smith, A. (1985). Language and cognitive processes from a developmental perspective. *Language and Cognitive Processes, 1*(1), 61-85.

Koster, C. (1993). *Errors in Anaphora Acquisition.* Unpublished Ph.D. Dissertation, Utrecht University, Utrecht.

Krifka, M. (1995). The Semantics and Pragmatics of Polarity Items. *Linguistic Analysis, 25*, 209-257.

Krifka, M. (2007). Approximate interpretation of number words: A case for strategic communication. In G. Bouma & I. Kräer & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 111-126). Amsterdam: Koninklijke Nederlandse Akademie van Wetenschapen.

Levinson, S. (2000). *Presumptive meaning: The theory of generalized conversational implicature*. Cambridge, Mass.: MIT Press.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492-527.

Mattausch, J. (2004). *On the Optimization & Grammaticalization of Anaphora.* Unpublished Ph.D. Thesis, Humboldt University, Berlin.

McCawley, J. D. (1978). Conversational implicature and the lexicon. In P. Cole (Ed.), *Syntax and Semantics 9: Pragmatics* (pp. 245-259). New York: Academic Press.

McKee, C. (1992). A Comparison of Pronouns and Anaphors in Italian and English Acquisition. *Language Acquisition, 2*(21-54).

Morgan, J. L. (1978). Two Types of Convention in Indirect Speech Acts. In P. Cole (Ed.), *Syntax and Semantics 9: Pragmatics* (pp. 261-280). New York: Academic Press.

Noveck, I. (2001). When children are more logical than adults. *Cognition, 78*, 165-188.

Noveck, I. A., & Sperber, D. (Eds.). (2005). *Experimental Pragmatics*. Houndmills, Basingstoke, Hampshire: Palgrave MacMillan.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year Olds' Difficulty with False Belief: The Case for a Conceptual Deficit. *British Journal of Developmental Psychology, 5*, 125-137.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. In J. L. McClelland & D. E. Rumelhart & the-PDP-Research-Group (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition 1*. Cambridge, MA: MIT Press.

Russell, B. (2006). Against grammatical computation of scalar implicatures. *Journal of Semantics, 23*, 361-382.

Sæbø, K. J. (2004). Optimal interpretations of permission sentences. In D. de Jongh & P. Dekker (Eds.), *Proceedings of the 5th Tbilisi Symposium on Language, Logic and Computation* (pp. 137-144). Amsterdam and Tbilisi.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy, 27*, 367 391.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing. *Psychological Review, 84*(1), 1-66.

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry, 27*, 720-731.

Smolensky, P., & Legendre, G. (2006). *The Harmonic Mind: From neural computation to optimality-theoretic grammar*. Cambridge, Mass.: MIT Press.

Soames, S. (1982). How presuppositions are inherited: A solution to the projection problem. *Linguistic Inquiry, 13*, 483-545.

Spenader, J., Smits, E.-J., & Hendriks, P. (2007). *Coherent discourse solves the Pronoun Interpretation Problem*.Unpublished manuscript, University of Groningen.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition, 57*, 31-95.

Sperber, D., & Wilson, D. (1986). *Relevance*. Oxford: Basil Blackwell.

Spohn, W. (1993). Wie kann die Theorie der Rationalität normativ und empirisch zugleich sein? In L. Eckensberger & U. Gähde (Eds.), *Ethik und Empirie. Zum Zusammenspiel von begrifflicher Analyse und erfahrungswissenschaftlicher Forschung in der Ethik* (pp. 151-196). Frankfurt a.M.: Suhrkamp.

Stalnaker, R. (2006). Saying and meaning, cheap talk and credibility. In A. Benz & G. Jäger & R. Van Rooij (Eds.), *Game Theory and Pragmatics* (pp. 82-101): Palgrave MacMillan.

Traugott, E. (1989). On the Rise of Epistemic Meanings in English: An Example of Subjectification in Semantic Change. *Language, 65*(1), 31-55.

Traugott, E., & Dasher, R. B. (2005). *Regularity in Semantic Change*. Cambridge: Cambridge University Press.

Unsworth, S. (2005). *Child L2, Adult L2, Child L1: Differences and Similarities.* Unpublished PhD Dissertation, Utrecht University.

Van Rooy, R. (2004). Signalling games select Horn strategies. *Linguistics and Philosophy, 27*, 493-527.

Wilson, D. (2003). Relevance and Lexical Pragmatics. *Italian Journal of Linguistics/Rivista di Linguistica, 15*(2), 273-291.

Wilson, D., & Matsui, T. (1998). Recent approaches to bridging: Truth, coherence, relevance. *UCL Working Papers in Linguistics, 10.*

Zeevat, H. (1999). Explaining presupposition triggers. *Proceedings of the Twelfth Amsterdam Colloquium, ILLC, Amsterdam*, 19-24.

Zeevat, H. (2000). The asymmetry of optimality theoretic syntax and semantics. *Journal of Semantics, 17*, 243-262.

Zeevat, H. (2002). Explaining presupposition triggers. In K. van Deemter & R. Kibble (Eds.), *Information Sharing* (pp. 61-87). Stanford: CSLI Publications.

Zeevat, H. (2007). *Optimal Interpretation as an Alternative to Gricean Pragmatics*.Unpublished manuscript, Universiteit van Amsterdam.

Zwarts, J. (2003). Lexical Competition: 'Round' in English and Dutch. In P. Dekker & R. van Rooy (Eds.), *Proceedings of the Fourteenth Amsterdam Colloquium* (pp. 229-234). Amsterdam: ILLC.

Zwarts, J. (2006). Om en rond: Een semantische vergelijking. *Nederlandse Taalkunde, 11*(2), 101-123.