

# EMBEDDED IMPLICATURES AND OPTIMALITY THEORETIC PRAGMATICS

REINHARD BLUTNER

UNIVERSITY OF AMSTERDAM

Blutner-AT-uva.nl

## Abstract

In a recent paper, Chierchia (2004) distinguishes global and local approaches to conversational implicatures and claims that several puzzles concerning implicatures in complex sentences can best be explained by a local approach. This conflicts with the Neo-Gricean view which is global in nature. I will argue that both approaches can coexist in optimality theoretic pragmatics where the proper place is assigned to the two approaches: a global theory describes the principal forces that direct communication – it has a diachronic dimension and allows a rational foundation of conversational implicatures; a local theory describes the actual, synchronic dimension – it explains how online, incremental interpretation of complex sentences is possible. The connection between the two views results from assuming that the results of global optimization fossilize into a local mechanism of utterance processing.

## 1 Introduction

In a recent paper, Chierchia (2004) distinguishes between global and local approaches to conversational implicatures. 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 (also called post-Gricean) 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 Gazdar (1979), Krifka (1995), Soames (1982), Blutner (1998), Sauerland (2004), Sæbø (2004; 2005), and Russell (2004); the local view is taken by Chierchia (2004), van Rooy (to appear), Levinson (2000), and Relevance Theory (e.g. Sperber & Wilson, 1986; Carston, 2002).

Whereas many globalists argue against the local view and many localists against the global view, I think 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. This way of distinguishing the two theories makes it possible to look for a systematic linking of the two perspectives. My suggestion is to take the idea of fossilization (or freezing/conventionalization/routinization) as a mediator between the two views. Though this suggestion is not much more than a speculative idea at the moment, I will argue that optimality theoretic (OT) pragmatics has the potential for contributing the linking theory.

The next section gives a concise introduction into OT pragmatics. It is explained how OT can account for both the diachronic and the synchronic perspective. Further, the idea of fossilization is explained with examples from Lexical Pragmatics. In Section 3 I explain how a global (Neo-Gricean) theory of conversational implicature can deal with several examples of embedded implicatures. Section 4 gives some arguments why a local account is required in order to describe the actual construction of implicatures, and it speculates about the role of fossilization. Section 5, finally, draws some general conclusions.

## **2 Optimality Theoretic Pragmatics**

OT can be seen as a general framework that systematizes the use of optimization methods in linguistics.<sup>1</sup> 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. 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 (Hendriks & de Hoop, 2001; de Hoop & de Swart, 2000).

Bidirectional optimization (Blutner, 1998, 2000) integrates the speaker and the hearer perspective into a simultaneous optimization procedure. In

---

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

pragmatics, this bidirectional view is motivated by a reduction of Grice’s maxims of conversation to two principles: the I/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. The Q-principle corresponds to the first part of Grice’s quantity maxim (*make your contribution as informative as required*), while it can be argued that the countervailing I/R-principle corresponds to the second part of the quantity maxim (*do not make your contribution more informative than is required*), the maxim of relation and possibly all the manner maxims. In a slightly different formulation, the I/R-principle seeks to select the most coherent interpretation and the Q-principle acts as a blocking mechanism which blocks all the outputs which can be grasped more economically by an alternative linguistic input. This formulation makes it quite clear that the Gricean framework can be conceived of as a bidirectional optimality framework which integrates the speaker and the hearer perspective. Whereas the I/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.

I will give a very schematic example in order to illustrate some characteristics of the bidirectional OT. Assume that 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 (marked) than the form  $f_2$  and that the interpretation  $m_1$  is less complex (marked) than the interpretation  $m_2$ . This is expressed by two markedness constraints F and M for forms and interpretations, respectively – 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).

Table 1: Markedness and bias constraints in a 2-forms  $\times$  2-interpretations design

	F	M	F→M	*F→*M	F→*M	F*→M
$\langle f_1, m_1 \rangle$					*	
$\langle f_1, m_2 \rangle$		*	*			
$\langle f_2, m_1 \rangle$	*			*		
$\langle f_2, m_2 \rangle$	*	*				*

Besides the markedness constraints 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 (Levinson’s (2000) **I**-constraint), the constraint \*F→\*M says that complex forms express complex interpretations

(Levinson's **M**-constraint<sup>2</sup>), and the two remaining bias constraints express the opposite restrictions. In the present case linking constraints can be seen as lexical stipulations that fix a form-interpretation relation in a memory (instance) based way.

In the so-called strong version of bidirectional OT, a form-interpretation pair  $\langle f, m \rangle$  is considered to be (strongly) optimal iff (I) no other pair  $\langle f, m' \rangle$  can be generated that satisfies the constraints better than  $\langle f, m \rangle$  and (Q) no other pair  $\langle f', m \rangle$  can be generated that satisfies the constraints better than  $\langle f, m \rangle$ . 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.<sup>3</sup> The optimal pairs are marked with the symbol ♯ in the diagram.

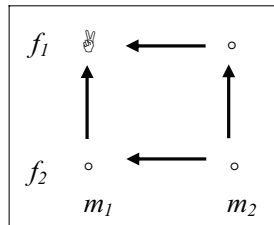


Figure 1: Diagram to illustrate strong bidirection

The scenario just mentioned describes the case of *total blocking* where some forms (e.g., \*furiousness, \*fallaciousness) 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. (1a)) 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,

<sup>2</sup> Levinson's M-principle should not be confused with the markedness constraint M introduced in Table 1.

<sup>3</sup> Dekker & van Rooy (2000), who introduced these diagrams, gave bidirectional OT a game theoretic interpretation where the optimal pairs can be characterized as so-called Nash Equilibria.

indirect causation. For example, (1b) could have been used appropriately when Black Bart caused the sheriff's gun to backfire by stuffing it with cotton.

- (1) 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 (1a),  $f_2$  the periphrastic form (1b),  $m_1$  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 bias-constraints  $F \rightarrow M$  and  $*F \rightarrow *M$  are higher ranked than the rest of the constraints. This can be depicted as in Figure 2a. Hence, strong bidirection can be taken as describing Horn's division of pragmatic labour when the appropriate linking constraints are dominating.

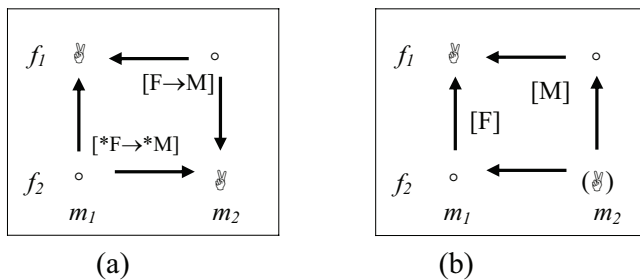


Figure 2: Two ways of describing Horn's division of pragmatic labour: (a) by assuming two dominant bias constraints; (b) by assuming markedness constraints and weak bidirection

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 bias constraints. Blutner (2000) develops a *weak* version of two-dimensional OT, according to which the two dimensions of optimization are mutually related: a form-interpretation pair

$\langle f, m \rangle$  is called super-optimal iff (I) no other *super-optimal* pair  $\langle f, m' \rangle$  can be generated that satisfies the constraints better than  $\langle f, m \rangle$  and (Q) 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 bias 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). The second case – involving 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 do not even fit the simplest requirements of a psychologically realistic model of online, incremental interpretation (Beaver & Lee, 2004; Zeevat, 2000). The proper understanding of weak bidirection relates to an off-line mechanism that is based on bidirectional learning (Benz, 2003; Blutner, Borra, Lentz, Uijlings, & Zevenhuijzen, 2002; Van Rooy, 2004). In these approaches the solution concept of weak bidirection is considered as a principle describing the direction of language change: super-optimal pairs are tentatively realized in language change. This relates to the view of Horn (1984) who considers the Q and the I principle as diametrically opposed forces in language change. This 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 in (1). 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 ( $f_i$ ) 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 bias constraints). According to this adaptation mechanism the expected 'utility' (how well they understand each other in the statistical mean) is improving 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 frequent to express in the population than the unmarked contents.<sup>4</sup>

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 bias constraints such as in Figure 2a. We will use the term *fossilization* for describing the relevant transfer.<sup>5</sup>

### **3 A global theory of embedded implicatures**

In OT pragmatics, a global theory of conversational implicatures is realized by using weak bidirection and systems of markedness constraints for forms and interpretations. In this section I will argue that the basic findings of conversational implicatures in complex sentences can be explained by this rational approach to communication.

---

<sup>4</sup> For more discussion of the role of frequencies in an evolutionary setting see Stalnaker (2006).

<sup>5</sup> Mattausch (2004) has implemented the idea of fossilization using stochastic OT. In that way he could explain the evolution of reflexive marking strategies in English and he was able to show how an optional and infrequent marking strategy like that of Old English could evolve into a pattern of obligatory structural marking like that attested in modern English.

### 3.1 Reciprocals, strength and relevance

A good candidate for a markedness constraint in the interpretive domain relates to the *strongest meaning hypothesis* (SMH). In its original formulation the constraint is used as a formal tool for analyzing the remarkable variation in the meaning of reciprocal expressions like *each other* or *one another* (e.g. Dalrymple, Kanazawa, Kim, Mchombo, & Peters, 1998). Consider for instance the following example:

- (2) a. The girls saw each other.
- b. The girls are standing on each other.

Sentence (2a) entails that every girl saw every other girl. This contrasts with sentence (2b) which obviously does not entail that each of the girls is standing on each of the others. The interpretation that is strongly preferred in these and similar cases is best described by the SMH given below:

SMH: A reciprocal sentence is interpreted as expressing the logically strongest candidate truth conditions which are not contradicted by known properties of the relation expressed by the reciprocal scope when restricted to the group argument.

Subsequent work has suggested to extend the application of the SMH for treating other phenomena with plurals (Winter, 2001), prepositions (Zwarts, 2003) and quantification (Blutner, Hendriks, & de Hoop, 2003).

Unfortunately, the SMH makes the wrong predictions in complex sentences such as

- (3) I doubt that the girls saw each other.

What is doubted in sentence (3) is the proposition that every girl saw every other girl. Of course, this doubt conforms to the strongest interpretation of the embedded sentence. Consequently, the complex proposition that is expressed by (3) relates to the *weakest* interpretation that is possible because of the negation element in the matrix sentence. It is justified already when some girl did not see any other girl. Hence, the global application of the SMH doesn't give the right result (we had to replace it by a *weakest meaning hypothesis* in downward entailing contexts!)

Advocates of the SMH have taken these arguments as showing that the principle is based on logical properties of lexical items and directly affects truth conditions. Because of its truth-conditional, nondefeasible nature, the SMH should be applied locally first. In a second step the projection of the generated

truth conditions to the complex sentence structure appears (using the means of compositional semantics). As a consequence, we have to give up the Gricean idea of conversational implicature in the context of reciprocals (and the other examples mentioned).

In Blutner (2006) I have argued, however, that a global, Gricean solution is possible if we replace the SMH by a principle of optimal relevance. Of course, it is essential to have a proper measure of relevance. Van Rooy (to appear) listed some candidate definitions he found in the linguistic, philosophical and statistical literature. For goal-oriented theories of relevance, but also for the entropy-based version it is essential that the value of relevance can be positive and negative. The maxim of optimal relevance then means maximizing the absolute amount of relevance.

Merin (1997) identified two crucial conditions for a proper theory of relevance, i.e. a theory of relevance that conforms to a compositional, linear mode of calculating the value of relevance for complex sentences:

- (4) a.  $\text{Rel}(A\&B) = \text{Rel}(A) + \text{Rel}(B)$  if propositions A and B are independent
- b.  $\text{Rel}(A) = -\text{Rel}(\neg A)$

Using a theory of relevance that satisfies these conditions (among them standard statistical relevance and Carnap's measure of relevance; cf. van Rooy, to appear), the Neo-Gricean approach can provide an explanation of the given examples. Of course we have to skip then the idea of *cancellability* as a general criterion for conversational implicatures.

An early example that justifies the idea that conversational implicatures are NOT necessarily cancellable is due to a classic paper by Sadock (1978):

Grice states explicitly that generalized conversational implicatures, those that have little to do with context, are cancellable. But is it not possible that some conversational implicatures are so little dependent on context that cancellation of them will result in something approaching invariable infelicity? In a paper in preparation, I argue that sentences of the form *almost P* only conversationally entail *not P*, contrary to the claim made by Karttunen and Peters (1979). The implicature is straightforwardly calculable and highly nondetachable but, unfortunately for my thesis, just about uncancellable. The sentence *Gertrude not only almost swam the English Channel, in fact she swam it is*, I admit, pretty strange. (Sadock 1978: 293)

In Section 3.3 I will come back to the cancellability issue.

### 3.2 Explicatures

In the relevance theoretic literature (Sperber & Wilson, 1986; Carston, 2002, 2003, 2004) the term *explicature* is used for pragmatic inferences that directly affect truth-conditions.

A key feature in the derivation of an explicature is that it may involve ‘free’ enrichment, that is, the incorporation of conceptual material that is wholly pragmatically inferred, on the basis of considerations of rational communicative behaviour, as these are conceived of on the relevance-theoretic account of human cognitive functioning (Carston, 2003: 819).

This is one of the standard examples:

- (5) a. John had a drink  $\rightsquigarrow$  John had an alcoholic drink
- b. I doubt that John had a drink  $\rightsquigarrow$  I doubt that John had an alcoholic drink

Other examples of free enrichment relate to domain restrictions in the case of quantification. In the following example the explicature depends on the context – assume the sentence is uttered in a typical party situation:

- (6) a. Everyone left early ( $\rightsquigarrow$  everyone at the party left early)
- b. Either everyone left early or the ones who stayed on are in the garden

Also mereonomic restrictions can be seen as not fully specified by the underlying semantics and thus constituting a task for free enrichment:

- (7) a. This apple is red ( $\rightsquigarrow$  the outside of the apple is red)
- b. I doubt that the apple is red

A straightforward observation is that all these examples are based on *I/R-implicatures* according to the Neo-Gricean classification. Hence, in order to give an explanation of the projection properties it is essential to have a proper measure of relevance as suggested above – a measure which hopefully can be extended to other complex forms than those constructed by negation and conjunction.

### 3.3 Scalar implicatures

Blutner (1998) proposed an approach to "scalar implicatures" that has some advantages over the traditional approach based on Horn-scales (see Gazdar, 1979). For example it solves a famous puzzle given by James D. McCawley. In the exercise part of his logic book McCawley (1993: 324) points out that the derivation of the exclusive interpretation by means of Horn-scales breaks down as soon as we consider disjunctions having more than two arguments. For example, from a disjunctive sentence of the form *John or Paul or Ede is sick* we can conclude that only one of the three is sick. However, the traditional approach predicts that not all the disjuncts can be true, which is too weak. The solution was to admit a whole lattice of alternative expressions constructed by the AND operator in order to block all interpretations with more than one individual sick.

As discussed in Blutner (2006), the global solution also works in cases like (8a) where the implicatures are (8b-c):

- (8)
- a. Someone is sick
  - b. The speaker does not know who is sick
  - c. The speaker knows (exactly) one individual is sick (in a given set of individuals)

The analysis is based on three general assumptions: (i) a Neo-Gricean theory of scalar implicatures based on a global blocking mechanism; (ii) Soames' (1982) reconsideration of the epistemic status of scalar implicatures paired with a default mechanism of neg-raising (Horn, 1989); (iii) a linear theory of relevance as suggested before. Related proposals are due to recent suggestions by Sauerland (2004) and Russell (2004).

It is not difficult to see how to analyse the projection behaviour of scalar implicatures via our global theory of implicature projection. I start with one of Carston's (2002) examples:

- (9)
- a. Mary lives somewhere in the south of France
  - b. Speaker does not know where in the south of France Mary resides
  - c. If Mary lives somewhere in the south of France, then I do not know where

Obviously, uttering (9a) implicates the proposition (9b). The derivation of this implicature is analogous to the derivation of (8b). However, the implicature does not locally arise in the antecedent of a conditional such as in (9c). If it would arise, then the whole sentence (9c) would be a tautology, but it is not. The

explanation in the present Neo-Gricean framework is obvious: the expression alternatives to (9c) have to be logically stronger than (9c) itself. Because the weak quantifier *somewhere in the south of France* occurs in the antecedent of a conditional, replacing it by concrete locations results in a weaker expression that does not count as an expression alternative. Hence, the implicature does not arise.

Chierchia (2004) discussed many other examples with scalar implicatures and concluded that only a local theory can account for the observed phenomena. However, Sauerland (2004) and Russell (2004) have shown that a global Gricean theory is appropriate to account for each of the implicatures Chierchia identified. In agreement with these authors we can conclude that a global account is possible for the treatment of scalar implicatures.

Scalar implicatures are normally considered to be cancellable. However, it has often been remarked that cancellability is difficult to distinguish from *clarification* (cf. Burton-Roberts, 2005) and also from *contextual change*. With regard to the latter point, van Kuppevelt (1996) has carefully argued that scalar implicatures are topic-dependent, i.e. they are dependent on the question being asked in a particular conversational setting. Consider the following example as discussed by van Rooy (to appear):

- (10) a. Question: Who has 2 children?
- b. Answer: John has 2 children
- c. John doesn't have more than 2 children

In this case, the implicature (10c) does not even arise. This is different from the following situation where the question is focussing on the number of children:

- (11) a. Question: How many children does John have?
- b. Answer: John has 2 children
- c. John doesn't have more than 2 children

In this case the implicature (c) arises; however, it cannot be cancelled. Van Kuppevelt (1996) argues that the 'phenomenon of cancellation' is in fact an effect of contextual change. In this sense scalar implicatures are *particularized* conversational implicatures. Obviously, the topic-dependency of scalar implicatures is not restricted to numerals but also holds in connection with the Q-implicature triggered by 'or' (cf. Van Rooy, to appear). The consequence of this finding is that cancellability cannot longer count as a criterion for identifying conversational implicatures.

### 3.4 Free choice interpretation

In a recent paper, Kjell Johan Sæbø (Sæbø, 2004) has developed a natural pragmatic solution to the long-standing problem of Free Choice Permission. In his theory the free choice assumption (12b) that is communicated by uttering sentences like (12a) comes out as a conversational implicature.

- (12) a. You may take an apple or a pear.  
 b. You may take an apple and you may take a pear.

Using the framework of Bidirectional OT his solution conforms to the global approach to conversational implicature. However, examples like (13) proposed by Kamp (1973) argue against seeing this inference as a purely conversational implicature.

- (13) Usually you may only take an apple. So, if you may take an apple or a pear, you should bloody well be pleased.

In cases like (13) the free choice inference projects like an explicature. The disjunctive permission is the antecedent of a conditional, and the strong interpretation of it rather serves to weaken the assertion than to strengthen it. Sæbø (2004) has interpreted this behaviour by considering the implicated assumption as constituted by an *utterance report* which is used in the embedded sentence ('they tell you that you may take an apple or a pear'). He points out that the strong interpretation is not always available in embedded positions and gives the following examples.

- (14) a. I hope you may take an apple or a pear; then, you won't feel so hungry.  
 b. If you may take an apple or a pear, you must take a pear.  
 c. If they tell you that you may take an apple or a pear, take a pear.

In example (14a), the free choice is not readily available, but in (14b) it is, at least if a paraphrase like (14c) is possible. Sæbø's treatment relates to Carston's (2004) analysis that also refers to different uses of the embedded material (Carston 2004; especially chapter 4). Carston distinguishes between material that is used *descriptively* or *meta-representationally*. In case of the meta-representational use the implicated assumption becomes part of the proposition expressed and projects like explicatures.

Generally, it can be concluded that the strategy of treating embedded material meta-representationally can be useful, but it still leaves us with the problem of explaining the exact projection behaviour of (ordinary) implicatures,

a task that cannot be solved by simply stipulating two modes of use (as RT does). Instead, implicatures in complex sentences can be explained best by the global, rationalist approach to communication where the introduction of meta-representational elements is modelled.

#### 4 Fossilization and a local theory of embedded implicatures

In the previous section I outlined a global approach to implicatures in complex sentences based on a Neo-Gricean framework of OT pragmatics. The approach rests on several assumptions that I will list here once more:

- (i) Merin's (1997) local theory of relevance.
- (ii) A Neo-Gricean theory of scalar implicatures based on a global blocking mechanism.
- (iii) Soames' reconsideration of the epistemic status of scalar implicatures paired with a default mechanism of neg-raising.

None of these assumptions has a stipulative character; rather each of them is motivated by independent evidence that has nothing to do with the projection behaviour of implicatures. While claiming that a global theory can explain complex implicatures, it is essential to state that a global theory cannot count as an actual mechanism of language performance, since it doesn't conform to the principles of online, incremental interpretation. Rather, a global account describes the general forces that direct communication. It has a diachronic dimension. In order to get a synchronic system which describes the actual pragmatic inferences, the idea of *fossilization* has been proposed (Blutner, 2006). A theory of fossilization describes how pragmatic inferences become automatized and form part of an efficient cognitive system that makes fast online processing possible. The theory conforms to a memory/instance theory of automatization (cf. Logan, 1988).

The presumption of fossilization can be seen as a theory that realizes Dawkins' (1983) idea of *memic selection*. This idea conforms to the "universal Darwinist" claim (Dennett, 1995) that the methodology of evolutionary theory is applicable whenever any dynamical system exhibits (random) variation, selection among variants, and thus differential inheritance. Related proposals are Steels' recruitment theory of cultural evolution (e.g., Steels, 1998) and Kirby's paradigm of iterated learning (e.g., Kirby, 2000).

OT is a system of knowledge representation that invites for the development of the evolutionary perspective because the manipulation of the different rankings of a given system of constraints is a powerful but computationally simple task. It has been applied to the area of lexical pragmatics, especially in order to explain the phenomenon of broadening and strengthening in connection

with the prepositions *om* and *rond* in Dutch (Zwarts, 2005). In a related paper (Blutner, 2006) I proposed to apply the theory to phenomena outside the realm of lexical pragmatics.

Though real simulation results are missing at the moment there are some psychological implications of the new perspective of fossilization. Recent data of Noveck's experimental pragmatics group (cf. Noveck, 2005) suggests that children are sometimes more logical than adults. In one of their experiments they presented children and adults with sentences such as (15a) where a relatively weak term is used in scenarios where a stronger term is justified. From a logical point of view (15a) is obviously true. We know that elephants in general have trunks, from which it logically follows that (at least) some of them do. Things are changing if we take pragmatics into account. Understood as carrying a scalar implicature, a sentence like (15a) is not true but false. In order to make the situation even more perplex for the subjects the experimental items like (15a) were presented along with control items that are patently true or patently false, like in (15b) and (15c) respectively:

- (15) a. Some elephants have trunks.
- b. Some houses have bricks.
- c. Some crows have radios.

Surprisingly, younger children are typically more likely than adults to find the experimental utterance acceptable. One possibility to interpret this data is to assume that children are pragmatically delayed at young ages. From the fossilization perspective, it can be claimed that scalar inferences become automatic with age and that the experimental results are simply revealing how such inference-making matures.

This way of explaining the basic empirical finding contrasts with the view of Relevance Theory that 'would suggest that children and adults use the same comprehension mechanisms but that greater cognitive resources are available for adults, which in turn encourages them to draw out more pragmatic inferences' (Noveck, 2005). Noveck's data seem to confirm Relevance Theory, especially the data that show a link between scalar-inference production and task complexity. However, the present view of fossilization does not necessarily conflict with these findings since memory-based automatization does not mean that the task complexity cannot have any influence.

## **5 Conclusions**

As with global and local accounts of lexical pragmatics, the present investigation sees global and local accounts of embedded implicatures as complementary: the global account has a diachronic dimension and allows a

rational foundation of conversational implicatures; the local account describes the actual, synchronic dimension – it explains how online, incremental interpretation of complex sentences is possible. The connection between the two views results from assuming that the results of global optimization fossilize into a local mechanism of utterance processing. A theory of fossilization has important consequences for overcoming the predominant, synchronic view of language, for assimilating language competence and language performance, and for getting a better understanding of the division of labour between semantics and pragmatics. At present, we do not have a real theory of fossilization, but it appears that OT pragmatics has the appropriate tools in order to formulate such a theory.

### Acknowledgements

My interest in the problem of implicatures in complex sentences was initiated by Kjell Johan's work about the optimal interpretations of permission sentences, and I began to see the relevance of his treatment for a general theory of implicatures in complex sentences. Hence, my first thanks go to Kjell Johan for his initiative. For valuable hints, discussion and support I have to thank Jay Atlas, Stefan Blutner, Noel Burton-Roberts, Torgrim Solstad, and Henk Zeevat.

### References

- Beaver, D., and H. Lee (2004): 'Input-output mismatches in OT', in R. Blutner & H. Zeevat (eds.): *Optimality Theory and Pragmatics*, Houndmills, Basingstoke, Hampshire: Macmillan/Palgrave.
- 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.
- Blutner, R. (1998): 'Lexical pragmatics', *Journal of Semantics* **15**, pp. 115-162.
- Blutner, R. (2000): 'Some aspects of optimality in natural language interpretation', *Journal of Semantics* **17**, pp. 189-216.
- Blutner, R. (2006): 'Optimality Theoretic Pragmatics and the Explicature/Implicature Distinction', in N. Burton-Roberts (ed.): *Advances in Pragmatics*, Houndmills, Basingstoke, Hampshire: Macmillan/Palgrave.
- Blutner, R., E. Borra, T. Lentz, A. Uijlings, and R. Zevenhuijzen (2002): 'Signalling games: Hoe evolutie optimale strategieën selecteert', *Handelingen van de 24ste Nederlands-Vlaamse Filosofiedag*, Amsterdam: Universiteit van Amsterdam.
- Blutner, R., H. de Hoop, and P. Hendriks (2005): *Optimal Communication*, CSLI.

- Blutner, R., P. Hendriks, and H. de Hoop (2003): 'A new hypothesis on compositionality', in P. P. Slezak (ed.): *Proceedings of the Joint International Conference on Cognitive Science*, Sydney: ICCS/ASCS.
- Blutner, R., and H. Zeevat (eds.) (2004): *Optimality Theory and Pragmatics*, Houndmills, Basingstoke, Hampshire: Palgrave/Macmillan.
- Burton-Roberts, N. (2005): 'Robyn Carston on semantics, pragmatics and 'encoding'', *Journal of Linguistics* **41**, pp. 389-407.
- Carston, R. (2002): *Thoughts and Utterances: The Pragmatics of Explicit Communication*, Oxford: Blackwell.
- Carston, R. (2003): 'Explicature and semantics', in S. David & B. Gillon (eds.), *Semantics: A Reader*, pp. 817-845, Oxford: Oxford University Press.
- Carston, R. (2004): 'Relevance theory and the saying/implicating distinction', in L. Horn & G. Ward (eds.), *Handbook of Pragmatics*, pp. 633-656, Oxford: Blackwell.
- 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.
- Dalrymple, M., M. Kanazawa, Y. Kim, S. Mchombo, and S. Peters (1998): 'Reciprocal expressions and the concept of reciprocity', *Linguistics and Philosophy* **21**, pp.159-210.
- Dawkins, R. (1983): *The extended phenotype*, Oxford: Oxford University Press.
- de Hoop, H., and H. de Swart (2000): 'Temporal adjunct clauses in optimality theory', *Rivista di Linguistica* **12**(1), pp. 107-127.
- Dekker, P., and R. van Rooy (2000): 'Bi-directional optimality theory: An application of game theory', *Journal of Semantics* **17**, pp. 217-242.
- Dennett, D. (1995): *Darwin's dangerous idea*, New York: Simon & Schuster.
- Gazdar, G. (1979): *Pragmatics*, New York: Academic Press.
- Hendriks, P., and H. de Hoop (2001): 'Optimality theoretic semantics', *Linguistics and Philosophy* **24**, pp. 1-32.
- 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.
- Jäger, G. (2002): 'Some notes on the formal properties of bidirectional optimality theory', *Journal of Logic, Language and Information* **11**, pp. 427-451.
- Kamp, H. (1973): 'Free choice permission', *Proceedings of the Aristotelian Society, New Series* **74**, pp. 57-74.
- Kirby, S. (2000): 'Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners', in C. Knight, M. Studdert-Kennedy & J. R. Hurford (eds.): *The Evolutionary Emergence of*

- Language: Social Function and the Origins of Linguistic Form*, pp. 303-323, Cambridge: Cambridge University Press.
- Krifka, M. (1995): 'The Semantics and Pragmatics of Polarity Items', *Linguistic Analysis* **25**, pp. 209-257.
- 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**, pp. 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.
- McCawley, J. D. (1993): *Everything that Linguists have Always Wanted to Know about Logic*, 2nd edition, Chicago, IL: University of Chicago Press.
- Merin, A. (1997): 'If all our arguments had to be conclusive, there would be few of them', *Arbeitspapiere SFB 340*, Stuttgart.
- Noveck, I. A. (2005): 'Pragmatic inferences related to logical terms', in I. A. Noveck & D. Sperber (eds.): *Experimental Pragmatics*, Houndmills, Basingstoke, Hampshire: Palgrave MacMillan.
- Russell, B. (2004): *Against grammatical computation of scalar implicatures*, Brown University, Department of Cognitive and Linguistic Sciences.
- Sadock, J. M. (1978): 'On testing for conversational implicature', in P. Cole (ed.), *Syntax and Semantics, Volume 9: Pragmatics*, pp. 281-297, Academic Press.
- 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.
- Sæbø, K. J. (2005): *Focus Interpretation in Thetic Statements: Alternative Semantics and OT Pragmatics*, ms. University of Oslo.
- Sauerland, U. (2004): 'Scalar implicatures in complex sentences', *Linguistics and Philosophy* **27**, pp. 367-391.
- Smolensky, P., and G. Legendre, (2005): *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**, pp. 483-545.
- Sperber, D., and D. Wilson (1986): *Relevance*. Oxford: Basil Blackwell.
- Stalnaker, R. (2006): 'Saying and meaning, cheap talk and credibility', in A. Benz, G. Jäger & R. van Rooy (eds.): *Game Theory and Pragmatics*, pp. 82-101, Palgrave MacMillan.
- Steels, L. (1998): 'The origins of syntax in visually grounded robotic agents', *Artificial Intelligence* **103**, pp. 133-156.

- van Kuppevelt, J. (1996): 'Inferring from Topics: Scalar Implicature as Topic-Dependent Inferences', *Linguistics and Philosophy* **19**, pp. 555-598.
- Van Rooy, R. (2004): 'Signalling games select Horn strategies', *Linguistics and Philosophy* **27**, pp. 493-527.
- Van Rooy, R. (to appear): 'Relevance of complex sentence', *Mind and Matter*.
- Winter, Y. (2001): 'Plural predication and the Strongest Meaning Hypothesis', *Journal of Semantics* **18**, pp. 333-365.
- Zeevat, H. (2000): 'The asymmetry of optimality theoretic syntax and semantics', *Journal of Semantics* **17**, pp. 243-262.
- 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. (2005): *Om en rond: Een semantische vergelijking*, Radboud, University Nijmegen.