Optimality Theoretic Pragmatics and the Explicature/Implicature Distinction

Reinhard Blutner
University of Amsterdam

Abstract
Optimality Theoretic Pragmatics is a (partly) formalized theory that conforms to a dynamic neo-Gricean approach. It assumes one phase of the updating process that involves the application of the so-called Q- and I/R-principles. Critics of the theory have maintained that such an approach does not discriminate between processes where apparent conversational implicatures enter into propositional content from processes where conversational implicatures supplement the propositional content without becoming part of it. Hence, it does not account for the Relevance-Theoretic distinction between explicatures and implicatures. In the present paper I discuss the possibility for reconstructing the distinction in Optimality Theoretic Pragmatics. After careful consideration of recent empirical observations on implicatures in complex sentences the conclusion is drawn that the distinction should not be stipulated by referring to separate principles of the cognitive architecture (neither by stipulating different *modes* of interpretation nor by assuming distinct *phases* of processing). Instead, the distinction seems to be a consequence of a global optimization mechanism the results of which can be frozen into a local projection mechanism that conforms to the principles of incremental interpretation.

1 Introduction

Communication is not a simple matter of *coding* and *decoding* as certain Cartesian theories of language have claimed. Relevance Theory (RT) carefully argues that inference is the basis of all communication, and of all aspects of linguistic communication (Sperber & Wilson 1986). Such inferences conform to certain expectations that are created by communication. The representatives of RT are followers of Grice in that they stress that these expectations play a key role in utterance interpretation and are therefore constitutive of the whole interpretation process. Unlike Grice, however, they don't postulate conversational "maxims" for providing the standards of rational discourse. Instead, they claim that interpretation is primarily a cognitive phenomenon which depends on how humans process information. The contract-like dimension of human communication is thus external to the interpretation mechanism and is not seen as a consequence of the nature of human cognition.

The identification of explicatures and implicatures as two aspects of the pragmatic dimension of natural language comprehension is an important insight of RT (Sperber & Wilson, 1986; Carston, 2002, 2004). The former are developments of the logical forms that are encoded by the sentence uttered and determine the propositional content of the utterance (truth-conditional pragmatics); the latter conform to expectations that exploit our encyclopaedic knowledge in order to derive more global, non truth-functional aspects of interpretation. The idea that the encoded semantic system of language is not fully propositional, i.e. it is not sufficient for determining propositional content (literal interpretation) crucially deviates from the view of Grice, who identifies the truth-conditional content of an utterance ("what is said") with its encoded sentence meaning. I will take this view of RT as basically correct.

In this paper I will take the RT stance of seeing NL interpretation as a cognitive phenomenon and thus considering the basic principles of communication as a consequence of the nature of human cognition. However, I will argue that this view does not necessarily conflict with the idea that pragmatic inferences[1] are the product of rational behaviour between

[1] I use this term collectively for implicatures and explicatures.

cooperative interlocutors. Going back to Zipf (1949), I will show that global, rationalist principles of communication conform to a *diachronic* view of language – describing the forces that direct language change. On the other hand, it is obvious that a cognitive theory has to account for the incremental, automatic mechanism of utterance interpretation, and thus requires the analysis of the actual realization of explicatures/implicatures. There is a connection between the diachronic and the synchronic view, between the global forces that generally direct communication and the actual, automatized inferences that evolve from these forces. In the following I will use the term *fossilization* in order to refer to this relationship. A theory of fossilization describes how pragmatic inferences become automatized and form part of an efficient cognitive system that makes fast online interpretation of utterances possible.

Accepting the dualism between a diachronic view and a synchronic view invites us to see the synchronic account as informed by the diachronic account. We should try to explain particular synchronic patterns of behaviour by the pragmatic forces that drive language change. If the distinction between implicatures and explicatures is a crucial distinction in actual language interpretation, then it is important to ask if this distinction can be grounded in these fundamental diachronic forces. I think there is a positive answer to this question: the global, rational view behind these diachronic forces can explain which pragmatic inferences behave like explicatures or implicatures, respectively.

This view contrasts with the view of RT, which rests on the stipulation of two distinct cognitive strategies/styles: one for the calculation of explicatures and the other for the calculation of implicatures. Unfortunately, the stipulation does not *explain* the fundamental distinction, and RT does not give any hints how to overcome this explanatory problem. Since explicatures are derived by pragmatic inference as much as implicatures, the discrimination cannot be derived from the nature of the inferential mechanism.

In this article I will propose a rational foundation of the distinction based on basic principles of cultural evolution (a Zipfean/Neo-Gricean theory of balancing pragmatic forces). However, this is not enough from the perspective of Cognitive Science. It further needs a theory that explains the actual interpretation process – an online account of processing explicatures and implicatures. The theory of fossilization is proposed for filling in this gap.

The theoretical framework I adopt is optimality-theoretic pragmatics (Blutner, 2000; Hendriks & de Hoop, 2001; Blutner & Zeevat, 2004; Blutner, de Hoop, & Hendriks, 2005). I will show that this framework can be used both for giving an appropriate reconstruction of the implicature/explicature distinction and also for developing the idea of fossilization. Unfortunately, the important phenomenological distinction between explicature and implicature was completely ignored in the early papers on OT pragmatics and also later the importance of the distinction was marginalized (but see the discussion in Blutner et al., 2005). The main reason for this weakness was that the proponents of OT pragmatics seldom considered *complex* sentences in calculating invited inferences but were mainly concerned with relatively simple sentences. As a matter of fact, the different behaviour of implicatures and explicatures can best be studied in the context of complex sentences. Hence, an examination of pragmatic inferences in complex sentences is essential. Gazdar (1979) pioneered this, and recently, Chierchia (2004) published a paper that has revitalized the investigation of invited inferences in complex sentences.

In the next section, I will examine the RT distinction between implicatures and explicatures, and I will redefine the distinction in order to find a more adequate basis for theoretical analysis. Needless to say, I don't intend to violate the *spirit* of RT with this clarification. I think my implementation of the distinction fits better with the examples provided by the proponents of RT themselves for illustrating the distinction. Section 3 gives a concise introduction into OT pragmatics. Using ideas of Mattausch (2004) and others, I will further show what a model of fossilization might look like. In Section 4 I investigate some examples that provide the empirical basis for my proposed distinction between implicatures

and explicatures, and I will sketch how my Neo-Gricean view can account for the explicature/implicature distinction. In Section 5, finally, I will discuss the consequences of the proposed theory for experimental pragmatics and I will draw some general conclusions.

2 The distinction between explicatures and implicatures

It is not a simple task to give a definition of the explicature/implicature distinction. The school of 'radical pragmatics' (cf. Cole, 1981) does not make the distinction, and from the perspective of a basically Gricean mechanism of pragmatic strengthening, the distinction appears not to be significant. As mentioned already, for RT the distinction is essential because RT crucially departs from Grice in that it sees linguistically encoded semantic information as not sufficient for determining the propositional content of an utterance. This relates to RT's underdeterminacy thesis, which gives the pragmatic component a much greater role in deriving communicated assumptions. Hence, it is not surprising that RT comes up with a new classification of pragmatically communicated assumptions. In the words of Carston (2003):

> *There are two types of communicated assumptions on the relevance-theoretic account: explicatures and implicatures. An 'explicature' is a propositional form communicated by an utterance which is pragmatically constructed on the basis of the propositional schema or template (logical form) that the utterance encodes; its content is an amalgam of linguistically decoded material and pragmatically inferred material. An 'implicature' is any other propositional form communicated by an utterance; its content consists of wholly pragmatically inferred matter (see Sperber & Wilson 1986, 182). So the explicature/implicature distinction is a derivational distinction and, by definition, it arises only for verbal (or, more generally, code-based) ostensive communication.* (p. 9)

As Burton-Roberts (2005) points out, in RT, an implicature is defined negatively – as a communicated assumption that is not an explicature. This contrasts with the original definition of Sperber and Wilson (1986) who have a similar description for explicatures like Carston (2003) but see implicatures as assumptions constructed by "developing assumption schemas retrieved from encyclopaedic memory" (Sperber & Wilson 1986: 181). Whereas the early Sperber/Wilson definition contrasts the development of (underspecified) *logical forms* with the development of *encyclopaedic assumptions*, the position of Carston (2003) is to contrast the *development* of a linguistically encoded logical form with other ways of deriving communicated assumptions. The definition of Sperber and Wilson (1986) is clearer than the definition of Carston (2003) because it contrast theoretical concepts that are simpler to understand than the contrasts in Carston's definition. This has to do with the availability of cognitive-linguistic theories which explain the distinction between logical forms and encyclopaedic assumptions. However, there is – so far I can see – no theory that explains the distinction between communicated assumptions derived by development and communicated assumptions derived by other means. Unfortunately, the problem with the definition of Sperber and Wilson (1986) is that it conflicts with the intuitive classification of several *clear cases*.

(1) John had a drink ⇝ John had an alcoholic drink
(2) Some students wrote an essay ⇝ not all students wrote an essay

(1) is considered a clear instance of an explicature (free enrichment) though it includes reference to the encyclopaedia. Conversely, (2) is a clear instance of an implicature (scalar implicature) though it does not include reference to the encyclopaedia (instead it has a meta-

linguistic character – it requires reference to the lexicon for accessing quantifier alternatives in the language under discussion). Hence, we are in trouble with the earlier definition and are left with the second one. But the problem is that neither Carston nor anybody else offers a definition of 'development'. In the words of Burton-Roberts (2005): "'Development' is a black hole at the centre of the theory." (p. 397)

In cases like this, where there is no proper definition, it is a good strategy to look for some empirical tests that help to clarify the situation and which might be used to operationalize the distinction. Fortunately, there are such tests, and among the candidates that have been proposed are (i) the independency principle (Recanati, 1989), (ii) the scope embedding test (Recanati 1989, Carston 2002) and (iii) a test based on cancellability (e.g. Burton-Roberts 2005). I will not go into an explanation of the independency principle since it has been rejected as a useful test by most authors (cf. Carston 2002). The scope embedding test, however, has been considered useful by many authors. Here is a concise presentation of this criterion called scope principle by Recanati (1989):

> **Scope principle**: *A pragmatically determined aspect of meaning is part of what is said (and therefore, not a conversational implicature), if – and perhaps only if – it falls within the scope of logical operators such as negation and conditionals.* (Carston 2002: 191)

Obviously, this principle is related to Green's 'Embedded Implicature Hypothesis' (EIH):

> **EIH**: *If assertion of a sentence S conveys the implicatum that p with nearly universal regularity, then when S is embedded the content that is usually understood to be embedded for semantic purposes is the proposition S&p.* (Green, 1998: 77)

We can see an 'implicatum' that regularly satisfies EIH as an explicature and an 'implicatum' that has systematic violations of EIH as an implicature. Carston considers the scope principle and EIH as a helpful tool "though it should probably not be given the status of a principle" (Carston 2002: 195).[2]

One property that is crucially discussed in connection with the implicature/explicature distinction is cancellability. Carston argues that cancellability cannot be a criterion that distinguishes explicature from implicatures: 'it is pragmatic inference quite generally that is cancellable/defeasible' (Carston 2002: 138). Hence, both explicatures and implicatures are cancellable and this property cannot distinguish them. Burton-Roberts rejects the idea of explicatures that are cancellable because, in RT, explicatures are constitutive of the truth-conditional content of an utterance (satisfying EIH): 'Cancellable explicature, then is a logical impossibility in Carston's own terms' (Burton-Roberts 2005: 401). I think the argument is convincing, and this might suggest that we can take cancellability as the criterion we are looking for. Unfortunately, cancellation is not only difficult to distinguish from *clarification* (as discussed by Burton-Roberts) but 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.[3] Consider the following example as discussed by van Rooy (to appear):

---

[2] (Recanati, 1993) rejects the test because of metalinguistic negation. Indeed, examples like "I am not his daughter; he is my father" suggest that sometime some property other than propositional content is falling within the scope of negation or other logical operators. However, this does not really undermine the usefulness of the test because of intonational and other cues that indicate the metalinguistic use.

[3] Another example is due to a classic paper by Sadock (1978) and discussed in Blutner (1998): '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

(3) 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 (3c) does not even arise. This is different from the following situation where the question is focussing on the number of children:

(4) a. Question: How many children has John?
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' that is normally discussed in connection with scalar implicatures has nothing do with genuine cancellation; instead it has to do with 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 is ruled out as a criterion that distinguishes explicatures from implicatures. Hence we are left with EIH as the crucial test. This suggests we should take the different projection properties of explicatures and implicatures as defining the distinction.

Since explicatures are derived by pragmatic inference as much as implicatures, RT has to stipulate two distinct cognitive strategies/styles in order to explain the difference between explicatures and implicatures. One strategy/style explains the cognitive mechanism of *developing* a logical form, the other deals with other forms of deriving pragmatic assumptions. In contrast, the present view is Neo-Gricean in nature and tries to explain pragmatic inferences as the product of rational behaviour between cooperative interlocutors. I claim that this global account makes it possible to derive the different projection behaviour of explicatures and implicatures and gives, moreover, a detailed description of the embedding contexts where even implicatures project with nearly universal regularity. However, as already mentioned there is a problem with such a global account because it does not apply to the actual, online interpretation process. Therefore, we propose to add a theory of fossilization for filling in this gap. We will argue in the following section that this approach relates to a general theory of cultural learning.

3 Pragmatics in OT

In this section I will give a concise, but informal introduction into optimality theoretic pragmatics. For a detailed discussion the reader is referred to original literature (e.g. Blutner & Zeevat, 2004; Blutner et al., 2005). Not surprisingly, the idea of optimization was present in the pragmatic enterprise from the very beginning. Much more than in other linguistic fields, optimality scenarios are present in most lines of thinking: Zipf's (1949) balancing between effect and effort, the Gricean conversational maxims (Grice, 1975), Ducrot's argumentative view of language use (Ducrot, 1980), the principle of optimal relevance in Relevance theory (Sperber & Wilson, 1986). Interestingly, more than one optimization procedure is involved in some of these accounts. For instance, the Neo-Gricean framework assumes two

---

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)

countervailing optimization principles called Q and I principle (Atlas & Levinson, 1981; Horn, 1984, who writes R instead of I).

Optimality Theory (OT) can be seen as a general framework that systematizes the use of optimization methods in linguistics. 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, the input to this process of optimization is an underlying linguistic representation. The output is the form as it is expressed. In syntax, the input is an underlying logical form, and 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; Hoop & de Swart, 2000).

Bidirectional optimization integrates the speaker's and the hearer's 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 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 collects the second part of the quantity maxim (*do not make your contribution more informative than is required*), the maxim of relation and possibly all the manner maxims. In a slightly different formulation, the I/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 (Blutner 1998). 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.

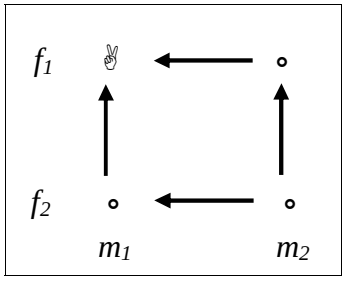
In the so-called strong version of bidirectional OT, a form-interpretation pair $\langle f, m\rangle$ is called (strongly) optimal iff (I/R) no other pair $\langle f, m'\rangle$ is generated that satisfies the constraints better than $\langle f, m\rangle$ and (Q) no other pair $\langle f', m\rangle$ is generated that satisfies the constraints better than $\langle f, m\rangle$. 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 set of interpretations with them, say $\{m_1, m_2\}$. 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_1. This is indicated in table (5).

(5)

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

From these differences of markedness the following ordering relation between form-meaning pairs can be derived as shown in (6). I'm using a graphical notation to indicate the preferences 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. The optimal pairs are marked with the symbol ✌ in the diagram.

(6)



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 further 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. (7a)) 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, (7b) could have been used appropriately when Black Bart caused the sheriff's gun to backfire by stuffing it with cotton.

(7) 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 (7a), f_2 the periphrastic form (7b), m_1 direct (stereotypical) causation and m_2 indirect causation.

Typical cases of total and 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 principal possibilities to avoid total blocking within the bidirectional OT framework. The first possibility is to formulate so-called bias constraints (Mattausch, 2004) and to find the appropriate ranking of the constraints such that partial blocking comes out.

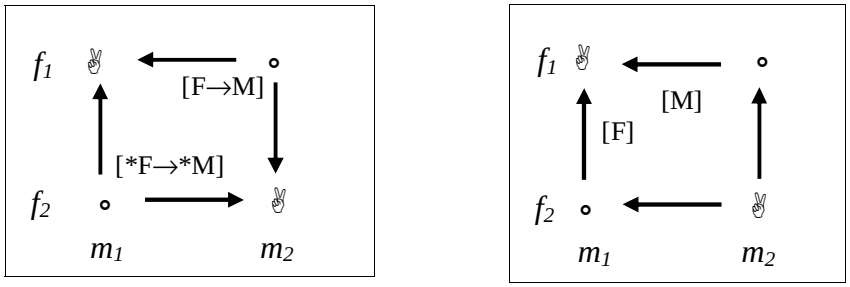
The table in (8) formulates four bias constraints besides the two markedness constraints F and M: the bias constraint F→M says that simple (unmarked) forms express simple interpretations (Levinson's I-constraint), the constraint *F→*M says that complex forms express complex interpretations (Levinson's (2000) M-constraint), and the two remaining bias constraints express the opposite restrictions.

(8)

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

Let's assume that the two bias-constraints F→M and *F→*M are higher ranked than the rest of the constraints. This can be depicted as in (9a). Hence, strong bidirection can be taken as describing Horn's division of pragmatic labour when the appropriate bias constraints are dominating.

(9) a. b.



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 <*f*, *m*> is called super-optimal iff (I/R) no other *super-optimal* pair <*f*, *m*'> is generated that satisfies the constraints better than <*f*, *m*> and (Q) no other *super-optimal* pair <*f* ', *m*> is generated that satisfies the constraints better than <*f*, *m*>. 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. Consider again the situation illustrated in (7), in the light of this weak version of bidirectional optimization.

It is not difficult to see that the *weak* version of bidirection can explain the effects of partial blocking without the stipulation of extra bias constraints; in particular, 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 $<f_1, m_2>$ and $<f_1, m_2>$ are *not* super-optimal.

Hence, they cannot block the pair <f_2, m_2> and it comes out as a new super-optimal pair. In this way, the weak version accounts for the Horn's d*ivision of pragmatic labour*. This is demonstrated in (9b). To stress this point again: to get this solution we only have to assume the markedness constraints F and M (alternately, we can assume that all the markedness constraints are higher ranked than the bias constraints).

The diagrams (9a) and (9b) describe the same set of solution pairs but the calculation of the solutions is completely different in the two cases. In the case of (9a) unidirectional optimization (either hearer or speaker perspective) is sufficient to calculate the solution pairs. I think this kind of OT system can be used to construct cognitively realistic models of online, incremental interpretation (Fanselow, Schlesewsky, Cavar, & Kliegl, 1999). But what about the status of weak bidirection (super-optimality) which is illustrated in (9b)? There are implementations of a recursive algorithm to calculate the super-optimal solutions (see various contributions in (Blutner & Zeevat, 2004)). Unfortunately, such a procedure doesn't even fit the simplest requirements of a psychologically realistic model of online, incremental interpretation because of its strictly non-local nature (Beaver & Lee, 2004; Zeevat, 2000).

Another problem is conceptual in nature and is raised by the existence of to notions of bidirectionality. What is the proper notion of optimization in natural language processing? The puzzle can be solved by relating weak bidirection to an off-line mechanism that is based on bidirectional learning. Benz (2003) worked out the formal details of such a theory. His theory is based on the idea that the speaker and hearer coordinate on form–meaning pairs which are most preferred from both perspectives. This theory predicts partial blocking as the result of an associative learning process where speaker and hearer preferences are coordinated. Other approaches ground weak bidirection likewise in repeated processes of bidirectional learning (cf. van Rooy (2004), Jäger (2004)). It is remarkable that in this research 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 principle and the I/R principle as diametrically opposed forces in inference strategies of language change. The basic idea goes back to Zipf (1949), and was reconsidered in van Rooy (2004) and Blutner et al. (2002). It conforms to the idea that synchronic structure is significantly informed by diachronic forces. Further, it respects Zeevat's (2000) acute criticism against super-optimality as describing an on-line mechanism.

For the sake of illustration, let's go back to the illustration in (9). Let's assume a population of agents who realize speaker- and hearer strategies based exclusively on the markedness constraints F and M. I.e., 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 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

(9a) where the two Levinsonian constaints I and 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 in the population than the unmarked contents.[4]

Hence, the important insight is that a system that is exclusively based on markedness constraints such as (9b) is evolutionarily related to a system based on highly ranked bias constraints such as (9a). We will use the term *fossilization* for describing the relevant transfer. The mentioned approaches of grounding weak bidirection in repeated processes of bidirectional learning can be seen as concrete realizations of fossilization. Recently, Mattausch (2000) has implemented the idea of fossilization using stochastic OT. In that way he was able to 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.

4 Embedded implicatures

There are two views of analyzing embedding implicatures. Following Chierchia (2004) I will call them the *global* and the *local view*, respectively. According to the global view one first computes the (plain) meaning of a complete sentence; then, taking into account the relevant alternatives, ones strengthens that meaning by adding in all the implicatures. 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 Gazdar (1979), Soames (1982), Blutner (1998), Sauerland (2004), and Russell (2004); the local view is taken by Chierchia (2004), van Rooy (to appear), Levinson (2000), and RT. 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 setting; local theories account for the shape of actual, online (synchronic) processing including the features of incremental interpretation. My suggestion is to take the proposal of fossilization as a mediator between the two views. However, at the moment this is not much more than an idea since concrete implementations of fossilization have applied for very simple examples only, predominantly in the domain of lexical pragmatics.

In the last section I introduced an OT implementation of the Neo-Gricean view. I will illustrate now how this global theory can account for the basic distinction between explicatures and implicatures. I will explain this by considering the projection behaviour of pragmatic inferences in complex sentences. What I will demonstrate is that explicatures regularly satisfy EIH whereas implicatures systematically violate EIH (predominantly in downward entailing contexts). Interestingly, the global theory does more than merely predict the basic distinction. Especially for scalar implicatures, it precisely predicts the projection behaviour's dependence on the surrounding context.

In Blutner (1998) I have proposed an approach to "scalar implicatures" that has some advantages over the traditional approach based on Horn-scales (cf. Gazdar 1979). For example it solves a famous puzzle given by John 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*

[4] There are examples where this condition is violated. For example, Zwarts (2005) considered the case of *om* and *rond* in Dutch where the unmarked term om refers to some strengthening of the frequent de tour interpretation, and the term rond refers to some weakening of the less frequent circle interpretation. The principle of strongest interpretation – taken as a markedness principle –, however predicts the opposite pattern.

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 then one individual sick.

The global solution also works in cases like (10a) where the implicatures are (10b&c)

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

Blocking sentences are of the form *i is sick* in this case, where *i* is the name of an individual. As van Rooy points out a Gazdarian analysis of scalar implicatures predicts that *i is not sick*, for any individual *i*. And this results in an *inconsistency* because we can draw that conclusion for each individual *i* and thus conclude that no one is sick. To resolve this problem Soames (1982) has proposed an alternative account based on a careful consideration of the quality maxim. He concludes that we should weaken the force with which the implicatures are generated. Instead of claiming that the speaker knows that the stronger proposition *i is sick* is not the case we should conclude only that the speaker does not know whether *i is sick* is the case. In that way we get the inference (10b). In order to get the implicature (10c) we can assume a neg-raising account of propositional attitudes (e.g. Horn, 1989). In the present context this conforms to an I-inference of the following kind:

(11) $\neg K\varphi \rightsquigarrow K\neg\varphi$, for any proposition φ [K is a belief operator]

Assuming this inference conforms to a default that is realized as long as no conflicts arise, we can derive the expected conclusion (10c). For a related proposal, see Sauerland (2004) and Russell (2004).

Now we are ready to analyse the projection behaviour of scalar implicatures in complex sentences. Compare the simple sentence (12a) and the complex sentence (12b):

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

Obviously, uttering (12a) implicates the proposition (12c). The derivation of this implicature is analogous to the derivation of (10b). However, the implicature does not locally arise in the antecedent of a conditional such as in (12b). Were it to arise, then the whole sentence (12b) would be a tautology, but it is not. Carston (2002: 194) uses this example to show that implicatures can violate EIH. The explanation in our Neo-Gricean framework is straightforward: the expression alternates to (12b) have to be logically stronger than (12b) itself. Because in (12b) the weak quantifier *somewhere in the south of France* occurs in the antecedent of a conditional (i.e. in a downward entailing context), replacing it by concrete locations results in a *weaker* expression. Since a weaker expression is not allowed as an expression alternative the implicature does not arise. We can conclude that scalar implicatures are indeed implicatures in the sense of RT rather than explicatures.

Another example confirms this conclusion.

(13) a. I believe that some students are waiting for me ⇝ I believe that some but not all students are waiting for me

b. I doubt that some students are waiting for me ↛ I doubt that some but not all students are waiting for me

c. ?Possibly all students are waiting for me. Hence, I doubt that some students are waiting for me

d. I doubt that some students are waiting for me ⇝ I believe that no students are waiting for me

In the upward entailing context (13a) the scalar implicature is realized as expected by the global account. In the downward entailing context (13b), however, the implicature does not project locally. Otherwise a discourse such as (13c) should be fine, but it is not. Rather, the implicature is as indicated in (13d), which relates to the logical negation of the quantifier. For a formal derivation the reader is referred to Russell (2004). It is essential that no extra stipulation is required besides those mentioned.

Chierchia discussed many other examples with scalar implicatures and concluded that only a local view can account for the observed phenomena. However, as Sauerland (2004) and Russell (2004) have shown, a Gricean theory can also account for each of the implicatures Chierchia identified. Hence, I conclude that a global account is possible for the treatment of scalar implicatures. The important question that arises now is whether a global account can also explain all the other pragmatic inferences that are discussed in the literature. In the rest of this chapter I will show how this account can explain examples typically treated as *explicatures* in the RT literature.

(14) a. John had a drink ⇝ John had an alcoholic drink

b I believe that John had a drink ⇝ I believe that John had an alcoholic drink

c. I doubt that John had a drink ⇝ I doubt that John had an alcoholic drink

This example shows that the inference of free enrichment in (14a) satisfies EIH both in upward entailing contexts (14b) and in downward entailing contexts (14c). Hence, it is an explicature in the sense of RT. The same projection behaviour is visible in the following examples when checking the pragmatic inferences given in parentheses:

Domain restrictions:

(15) a. Everyone left early (⇝ everyone at the party left early)

b. Either everyone left early or the ones who stayed on are in the garden

Meronomic restrictions:

(16) a. This apple is red (⇝ the outside of the apple is red)

b. I doubt that the apple is red

Reciprocals and plural predication

(17) a. The girls saw each other (⇝ every girl saw every other girl)

b. I doubt that the girls saw each other. No girl sees girl No. 5

(18) a. The cats see the dogs (⇝ every cat sees every dog)

b. I doubt that the cats see the dogs. No cat sees dog No. 3

(19) a. The cats are sitting in the baskets (⇝ every cat is sitting in one of the baskets)

b. I doubt that the cats are sitting in the baskets. Cat No. 1 is not sitting in any basket.

A first 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. Van Rooy (to appear) lists some candidate definitions 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.

Building on Merin (1997) van Rooy (to appear) identifies two crucial conditions for a local 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:

(20) a. Rel(A&B) = Rel(A) + Rel(B) if propositions A and B are independent
b. Rel(A) = –Rel(¬A)

Using such a local theory of relevance (which hopefully can be extended to other complex forms than those constructed by negation and conjunction), the Neo-Gricean approach can explain that the given examples exhibit *explicatures*, i.e. they satisfy EIH.

The explanation runs as follows. First, notice that unidirectional optimization (hearer perspective) is sufficient in the present case because blocking does not take place in the examples under discussion. Further, let's assume that GEN generates all possible enrichments of the logical form of the (complex) sentence; my favorite realization for such a mechanism of 'developing logical forms' is abduction (cf. Blutner, 1998). Finally, we assume that the evaluation component maximizes the amount of relevance (in absolute terms).

Assume an enrichment *m* of *LF*. By using a local enrichment mechanism it results that ¬*m* is an enrichment of ¬*LF*. Assuming the condition (20b), Rel(*m*) = -Rel(¬*m*), yields the following conclusion: if *m* is an *optimal* enrichment of *LF* then ¬*m* is an optimal enrichment of ¬*LF*. Hence, it can be concluded that EIH is inherited by negation, i.e. if a structure *S* satisfies EIH, then also ¬*S* satisfies it. Since *believe* and *doubt* are related by negation, it is obvious that in these contexts the I/R inferences qualify as explicatures. Assuming that Merin's linear pragmatics can be extended to other complex sentences, then it generally holds that all pragmatic inferences that count as I/R implicatures (Neo-Gricean terminology) are explicatures in the terminology of RT. I suggest this claim is empirically supported.

5 Consequences and conclusions

This article contains some speculation about the possibility of deriving the explicature/implicature distinction within a Neo-Gricean framework of optimality theoretic pragmatics. The speculations rest on several assumptions that I will list here once more:

(a) The nature of the distinction has to do with the embedded implicature hypothesis EIH: explicatures regularly satisfy EIH, implicatures regularly violate EIH in a definite class of contexts.
(b) A Neo-Gricean theory of scalar implicatures based on a global blocking mechanism.
(c) Soames' reconsideration of the epistemic status of scalar implicatures paired with a default mechanism of neg-raising.
(d) A local theory of relevance.

None of these assumptions has a stipulative character. With exception of the first assumption they are all motivated by independent evidence that has nothing to do with the explicature/implicature distinction.

However, while claiming that a global theory can explain the distinction, it is essential to state that a global theory cannot count as an online mechanism, since it doesn't conform to the principles of 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 (see Section 3). A theory of fossilization can be seen as describing how pragmatic inferences become automatized and form part of an efficient cognitive system that makes fast online processing possible. In this way, the theory conforms to a memory/instance theory of automatization (cf. Logan, 1988).

From the perspective of cultural evolution the presumption of fossilization relates to a theory that realizes Dawkins' (1983) idea of *memic selection*. This idea conforms to the "universal Darwinist" claim 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 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 in 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 the present paper I have proposed applying 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 presented by 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 (21) where a relatively weak term is used in scenarios where a stronger term is justified.

(21) Some elephants have trunks

Surprisingly, younger children are typically more likely than adults to find the utterance acceptable. One way of interpreting 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 contrasts with the suggestion made in RT '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).[5]

With regard to binding phenomena many researchers found a delayed principle B effect in comprehension but not in production. Children correctly interpret reflexives like adults from the age of 3;0 but they continue to perform poorly on the interpretation of pronouns even up

[5] Noveck's data seem to confirm RT, especially the data that show a link between scalar-inference production and task complexity. However, the present fossilization view does not necessarily conflict with these findings since memory-based automatization does not mean that the task complexity cannot have any influence (cf. Blutner & Sommer, 1988). Also the idea that context determines in many cases which pragmatic enrichment to make does not necessarily conflict with the fossilization view. This view assumes that fossilized pragmatic inferences implicatures can be context-dependent but they are not cancelable. This contrasts with Levinson's (2000) view, which clearly is refuted by these data.

to the age of 6;6. This contrasts with production data where even the youngest children use the pronoun to express a disjoint meaning while they use the reflexive to express a coreferential interpretation. This is puzzling because, usually, comprehension of a given form precedes production of this form.[6] In unpublished work it has been shown that fossilization theory can explain these data without further stipulation. The pragmatic inferences in this case are explicatures in the sense of RT, and it would be interesting to see how RT can account for these effects.

The present examination sees global and local accounts of analyzing embedding implicatures as complementary. (And the same holds for global and local accounts of lexical pragmatics). Hence, peaceful coexistence and even collaboration between globalists and localists is possible and desirable.

Acknowledgments

I am grateful to Jason Mattausch, Robert van Rooy, Torgrim Solstad and Henk Zeevat for valuable suggestions and comments on earlier versions of this paper. Special thanks to Noel Burton-Roberts for valuable comments, advice and painstaking editorial help.

References

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.

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

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*, 115-162.

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

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., & Sommer, R. (1988). Sentence Processing and Lexical Access: The Influence of the Focus-Identifying Task. *Journal of Memory and Language, 27*, 359-367.

Blutner, R., & Zeevat, H. (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*, 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.

[6] For a summary of the data and an interesting theoretical proposal cf. Hendriks & Spenader (2004).

Cole, P. (Ed.). (1981). *Radical pragmatics*.New York: Academic Press.
Dawkins, R. (1983). *The extended phenotype*.Oxford: Oxford University Press.
Ducrot, O. (1980). *Les Echelles argumentatives*.Paris: Minuit.
Fanselow, G., Schlesewsky, M., Cavar, D., & Kliegl, R. (1999). Optimal parsing, syntactic parsing preferences, and Optimality Theory.
Gazdar, G. (1979). *Pragmatics*.New York: Academic Press.
Green, M. (1998). Direct reference and implicature. *Philosophical Studies, 91*, 61-90.
Grice, P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics, 3: Speech Acts* (pp. 41-58). New York: Academic Press.
Hendriks, P., & de Hoop, H. (2001). Optimality theoretic semantics. *Linguistics and Philosophy, 24*, 1-32.
Hendriks, P., & Spenader, J. (2004). *When production precedes comprehension: An optimization approach to the acquisition of pronouns*.Unpublished manuscript, Groningen.
Hoop, H. d., & de Swart, H. (2000). Temporal adjunct clauses in optimality theory. *Rivista di Linguistica, 12*(1), 107-127.
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*, 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.
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.
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.
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, 101*.
Noveck, I. A. (2005). Pragmatic inferences related to logical terms. In I. A. Noveck & D. Sperber (Eds.), *Experimental Pragmatics*.Houndmills, Basingstoke, Hampshire: Palgrave MacMillan.
Recanati, F. (1989). The pragmatics of what is said. *Mind and Language, 4*, 294–328.
Recanati, F. (1993). *Direct Reference: From Language to Thought*.Oxford: Blackwell.
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.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy, 27*, 367– 391.
Soames, S. (1982). How presuppositions are inherited: A solution to the projection problem. *Linguistic Inquiry, 13*, 483-545.
Sperber, D., & Wilson, D. (1986). *Relevance*.Oxford: Basil Blackwell.
Steels, L. (1998). The origins of syntax in visually grounded robotic agents. *Artificial Intelligence, 103*, 133–156.
van Kuppevelt, J. (1996). Inferring from Topics: Scalar Implicature as Topic-Dependent Inferences. *Linguistics and Philosophy, 19*, 555-598.
Van Rooy, R. (2004). Signalling games select Horn strategies. *Linguistics and Philosophy, 27*, 493-527.
Van Rooy, R. (to appear). Relevance of complex sentence. *Mind and Matter*.
Zeevat, H. (2000). The asymmetry of optimality theoretic syntax and semantics. *Journal of Semantics, 17*, 243-262.
Zipf, G. K. (1949). *Human behavior and the principle of least effort*.Cambridge: Addison-Wesley.
Zwarts, J. (2005). Om en rond: Een semantische vergelijking:Radboud University Nijmegen.