

Pragmatics and the Lexicon

REINHARD BLUTNER
Humboldt University Berlin

This contribution investigates the interactions between the (mental) lexicon and pragmatics. It aims to give an overview about pragmatic phenomena that are connected with the semantic underspecification of lexical items. Cases in point are the pragmatics of adjectives, effects of negative strengthening, systematic polysemy, the distribution of lexical and productive causatives, blocking phenomena, the interpretation of compounds, and many phenomena presently discussed within the framework of Cognitive Semantics.

After emphasizing some important consequences of the traditional view of lexical semantics—the contrastive analysis of lexemes within the Katz-Fodor tradition of semantics—several phenomena are collected that seem to conflict with its theoretical settings. These phenomena are taken as arguments in favor of a particular account of the division of labor between lexical semantics and pragmatics. This account combines the idea of (radical) semantic underspecification in the lexicon with a theory of pragmatic strengthening, based on conversational implicatures. The basic pragmatic mechanism can be expressed within the framework of (bidirectional) optimality theory for interpretation. It is argued that this approach may provide a principled account of several of the lexical-pragmatic phenomena currently under discussion.

1. Introduction

In the view of Katz & Fodor (1963: 176) the scope of a language description covers the knowledge of a fluent speaker “about the structure of his language that enables him to use and understand its sentences”. The scope of a semantic theory is then the part of such a description not covered by a theory of syntax. There is a second aspect which Katz and Fodor make use of in order to bound the scope of semantics. This is the pragmatic aspect of language and it excludes from the description any ability to use and understand sentences that depends on the “setting” of the sentence. *SETTING*, according to Katz & Fodor (1963), can refer to previous discourse, socio-physical factors and any other use of “non-linguistic” knowledge. A nice demonstration of the essence of “non-linguistic” knowledge in the understanding of sentences was provided by psychologists in the 1970’s (e.g. Kintsch 1974). Let’s consider the following utterance:

- (1) The tones sounded impure because the hem was torn.

I suggest we do not really understand what this sentence means until we know that this sentence is about a bagpipe. It is evident that this difficulty is not due to our insufficient knowledge of English. The syntax involved is quite simple and there are no unknown words in the sentence. Instead, the difficulty is related to troubles in accessing the relevant conceptual setting. The idea of bagpiping is simply too unexpected to be derived in a quasi-

neutral utterance context. The example demonstrates that we have to distinguish carefully between the linguistic aspects of representing the (formal) meaning of sentences and the pragmatic aspects of utterance interpretation (speaker's meaning).

In this contribution I restrict myself to the semantics of lexical units and intend to explain the interaction of lexical meaning with pragmatics. Katz & Fodor (1963) already have stressed the point that a full account of lexical meaning has to include more information than that which allows one to discriminate the meanings of different words. In one of their examples they argue that *take back* is used in very different ways in the sentences (2a,b), although the relevant lexical entries are semantically unambiguous.

- (2) a. Should we take the lion back to the zoo?
b. Should we take the bus back to the zoo?

An obvious difference between these sentences is that the lion is the **object** taken back to the zoo in (2a), but the bus is the **instrument** that takes us back to the zoo in (2b). The problem for the pragmatic component of utterance interpretation is to explain the difference in terms of different conceptual settings, starting from a lexicon that doesn't discriminate the two occurrences of *take back* semantically and from a syntax that is completely parallel for the two sentences.¹

As another introductory example let's consider the perception verbs of English (cf. Sweetser 1990). If Saussure is right, there is an essential arbitrary component in the association of words or morphemes with what they mean. Consequently, the feature of arbitrariness could be taken at least as a sufficient condition for the presence of semantic information. It is certainly an arbitrary fact of English that *see* (rather than, say, *buy* or *smell*) refers to visual perception when it is part of the utterance (3a). Given this arbitrary association between a phonological word and its meaning, however, it is by no means arbitrary that *see* can also have an epistemic reading as in (3b).

- (3) a. I see the tree.
b. I see what you're getting at.

Moreover, it is not random that other sensory verbs such as *smell* or *taste* are not used to express an epistemic reading. Sweetser (1990) tries to give an explanation for such facts and insists that they have to do with conceptual organization. It is our knowledge about the inner world that implicates that vision and knowledge are highly related, in contrast to, say, smelling and knowledge or taste and knowledge, which are only weakly related for normal human beings. If this claim is correct, then the information that *see* may have an epistemic reading but *smell* and *taste* do not must no longer be stipulated semantically. Obviously, this can be formalized by language-independent, universal preferences. In specific languages, these preferences can be overridden – consider the “ingestive” verbs in languages of India, or even English “I can't swallow that” = ‘can't believe’. Instead, this information is pragmatic in nature, having to do with the utterance of words within a conceptual setting, and can be derived by means of some general mechanism of conceptual interpretation.

Considerations of this kind raise a standard puzzle for lexical semantics when we ask

how to separate the (mental) lexicon from the (mental) encyclopedia. How should we separate information about the meaning of words from information about the (supposed) reality associated with these words? Admittedly, it may be rather difficult to distinguish these two kinds of information. Tangible, theory-independent empirical tests simply don't exist. There are two principal possibilities for dealing with this situation. First, the distinction between the lexicon and the encyclopedia is said to be illusory (as has sometimes been suggested by proponents of Cognitive Semantics, e.g. Lakoff 1987). In this case all the relevant information has to be put into the lexicon. It will be argued in what follows that this view leads to a highly non-compositional account of meaning projection. The second possibility is to take the distinction as an important one. As a consequence, we are concerned with two different types of mechanisms:

- ◆ a mechanism that deals with the combinatorial aspects of meaning
- ◆ a pragmatic mechanism that deals with conceptual interpretation.

Once we have adopted such theoretical mechanisms, the problem of discriminating lexical semantic information from encyclopedic information need no longer look so hopeless, and we really may profit from a division of labor between semantics and pragmatics. It is the position of this contribution to argue in favor of the second option.

From a Gricean perspective, two different ideas of how to overcome the divergences between (formal) meaning and natural language interpretation come to mind. The first one uses conventional implicatures as an enlargement of the classical information entries. The second idea uses conversational implicatures as a method to overcome the divergences. While I believe that modern semantic theories (which usually are characterized as dynamic, epistemic, and non-monotonic) make the conception of conventional implicature superfluous as an addendum to the semantic component, I do not think the same is true of conversational implicature. In fact, in this paper I will argue that the proper use of conversational implicature will resolve some of the problems of lexical interpretation that remain unsolved otherwise.

The conceptual core of the theory I want to propose demands a straight formulation of conversational implicature. Paired with the idea of (radical) semantic underspecification in the lexicon and an appropriate representation of contextual and encyclopedic knowledge, this conception avoids unmotivated lexical ambiguities as well as the need for expansive re-interpretation and coercion mechanisms.

There are two basic aims of this paper. First, I want to demonstrate some general problems we are confronted with when trying to analyze the utterance of words within concrete conceptual and contextual settings and to go beyond the aspects of meaning typically investigated by a contrastive analysis of lexemes within the Katz-Fodor tradition of semantics. This may help to develop a sensitive feeling for what kind of problems may be approached by means of the division of labor between lexical semantics and pragmatics. Second, I would like to argue in favor of a particular account of the interaction between lexical semantics and pragmatics, one that combines the idea of (radical) semantic

underspecification in the lexicon with a theory of pragmatic strengthening (based on conversational implicature). It is illustrated that this view conforms with recent attempts to extend the framework of optimality theory (originally proposed by Prince & Smolensky 1993) for the purpose of natural language interpretation.

The organization of this paper is as follows. In the next section I will emphasize some important consequences of the traditional view of (lexical) semantics. In the third section some phenomena are collected that have a *prima facie* claim on the attention of linguists, and I will show that most of these phenomena conflict with the theoretical assumptions made by the traditional view. In the fourth section I introduce a particular way of combining (radical) semantic underspecification with a theory of pragmatic strengthening. Finally, the fifth section shows that this view can be expressed very naturally by using a (bidirectional) optimality theory of interpretation.

2. The standard view of (lexical) semantics

In this section I will remain neutral about what sort of thing a semantic value should be taken to be: an expression in some language of thought, a mental structure as applied in cognitive semantics or a model-theoretic construct. To be sure, there are important differences between conceptualistic accounts à la Katz & Fodor and realistic accounts as developed within model-theoretic semantics. These differences become visible, first at all, when it comes to substantiate the relationship between individual and social meaning (see Gärdenfors 1993). For the purpose of the present paper, however, the question of whether semantics is realistic or conceptualistic doesn't matter. In the following I will concentrate on some general features that can be ascribed to both accounts in their classical design. These features are not intended to completely characterize the family of theories representing the "standard view" in any sense. Rather, their selection is intended to emphasize several properties that may become problematic when a broader view of utterance meaning is taken. In sections 4 and 5, I will use these features for marking out the borderline between semantics and pragmatics.

2.1 Systematicity and compositionality

One nearly uncontroversial feature of our linguistic system is the SYSTEMATICITY OF LINGUISTIC COMPETENCE. According to Fodor & Pylyshyn (1988: 41-42), this feature refers to the fact that the ability to understand and produce some expressions is intrinsically connected to the speaker's ability to produce and understand other semantically related expressions. The classical solution to account for the systematicity of linguistic competence crucially makes use of the PRINCIPLE OF COMPOSITIONALITY. In its general form, tracing back at least to Frege (1892), this principle states the following:

- (4) The meaning of a complex expression is a function of the meanings of its parts and their syntactic mode of combination.

In an approximation that is sufficient for present purposes, the principle of

compositionality states that “a lexical item must make approximately the same semantic contribution to each expression in which it occurs” (Fodor & Pylyshyn 1988: 38). As a simple example consider adjective-noun combinations as *brown cow* and *black horse*. Let’s take absolute adjectives (such as *brown* and *black*) as one-place predicates. Moreover, non-relational nouns are considered as one-place predicates as well. Let’s assume further that the combinatorial semantic operation that corresponds to adjectival modification is the intersection operation. Fodor & Pylyshyn (1988) conclude that these assumptions may explain the feature of systematicity in the case of adjectival modification. For example, when a person is able to understand the expressions *brown cow* and *black horse*, then she should understand the expressions *brown horse* and *black cow* as well. Note that it is the use of the intersection operation that is involved in explaining the phenomenon, not compositionality per se. Nevertheless the principle of compositionality is an important guide that helps us to find specific solutions to the puzzle of systematicity.

Lexical semantics is concerned with the meanings of the smallest parts of linguistic expressions that are assumed to bear meaning. Assumptions about the meanings of lexical units are justified empirically only in so far as they make correct predictions about the meanings of larger constituents. Consequently, though the principle of compositionality clearly goes beyond the scope of lexical semantics, it is indispensable as a methodological instrument for lexical semantics. I state the principle of compositionality as the first feature characterizing the standard view of (lexical) semantics.

2.2 The monotonicity of inferential competence

The SYSTEMATICITY OF INFERENCE is another important feature of the standard view that was emphasized by Fodor & Pylyshyn (1988). Having in mind the logical vocabulary of natural language, the authors stress the common claim that these elements trigger systematic inferential competences. To be accurate, the systematicity feature of inference refers to the structure-sensitivity of the inferential relation. This contrasts with ASSOCIATIONS, which are not seen as structure-sensitive. As a standard example take the rule of SIMPLIFICATION in natural deduction, which is one of two rules connected with logical conjunction:

$$(5) \quad \frac{P \ \& \ Q}{\therefore P}$$

In a conditional proof, the constituent structure of the rule proves essential. In example (5) the constituent symbols P and Q function as place holders or variables, without having any intrinsic content. This trait allows different instantiations in one and the same actual proof. For instance, when starting with the premise $p \ \& \ q \ \& \ r$, we can infer p, q and r, inter alia, by applying the same rule (5) different times.

Why does the **content** of the premises not affect the inferences drawn? I claim that the answer to this question has to do with the more general idea of the MONOTONICITY OF

INFERENCE COMPETENCE. Tarski (1930, 1935) was the first to state the idea of monotonicity as one of three conditions that aimed to reflect the minimal requirements which a deductive inferential relation must fulfill if it is truly to be a logical relation. In informal terms, the condition says that old theorems remain valid when the system of axioms (definitions, meaning postulates, factual knowledge) has been augmented by adding some new axioms.²

For the sake of illustration, assume that the “content” of the elementary expressions p , q and r is partially described by some additional premises Π . If we add Π to the original expression $p \& q \& r$, then we expect that the “old” inferences p , q and r remain valid – due to the monotonicity of our inferential competence. Otherwise, we could not be sure that the old inferences survive, and the content of the constituent expressions p , q , r would affect the inferential potential of the logical conjunction – a rather absurd idea. Hence, the idea that logical inferences respect this pattern of monotonicity is so natural that it may appear to be unavoidable.

We conclude that the systematicity of inferential competence is intrinsically connected with the monotonicity restriction of the inferential relation. Without this restriction the systematicity of inference can become lost for the most part. However, what is good for mathematics must not necessarily conform to the laws of cognition in the general case. The monotonicity restriction is an empirical issue concerning our inferential competence. As such it has to be carefully checked.

2.3 The monotonicity of the lexical system

Another general characteristic of the standard view is connected with the idea of analyzing the meanings of lexical items as a complex of more primitive elements. The main motivation for such a COMPONENTIAL ANALYSIS is connected with the explanation of such semantic relations as antonymy, synonymy, and semantic entailment. If the meaning of a lexical item were not analyzable into components, the lexical system of grammar would have to simply enumerate the actually realized relations as independent facts. This procedure would be descriptively uneconomical. More important, it would miss the point that these facts are NOT independent from each other. The componential approach can be found both in theories of meaning in generative semantics (cf. Fodor 1977) and in model-theoretic based (especially Montagovian) semantic work (cf. Dowty 1979).

Defining the meaning of lexical items in terms of a repertoire of more primitive elements leads to a second order property which I will call the MONOTONICITY OF THE LEXICAL SYSTEM. In short, this monotonicity restriction refers to the fact that we can incrementally extend the lexical system (by adding some definitions for new lexical material) without influencing the content of elements already defined.

At first glance, the monotonicity of the lexical system looks quite natural as a constraint within formal semantics. Of course, it would be very surprising if the content of *...is a bachelor* were to change if the system learns what a spinster is (by acquiring the corresponding definition). Similarly, the meaning of *prime, even, odd (number)* should be independent of whether the system knows the meaning of *rational number* or *perfect number*³.

It should be stressed that it is not the idea of decomposition (definition) per se that leads

to the monotonicity feature of the lexical system. Instead, it is its classical treatment within a formal metalanguage that exhibits all features of a deductive system in the sense of Tarski. In this vein, the monotonicity of the lexical system can be seen as a specific realization of the more general aspect of the monotonicity of our inferential competence.

The following picture illustrates the difference between monotonic lexical systems and non-monotonic ones in a schematic way. The picture simplifies matters by identifying meanings with extensions (represented by Venn-diagrams). In the case of a monotonic system, the addition of a new predicate R doesn't change the extensions of the old predicates P and Q. However, the same doesn't hold in the case of a non-monotonic system. In this case we have **field-effects**: there seem to be attracting and repelling **forces** that shift the extensions of old predicates in a particular way when new lexical material comes into play.⁴

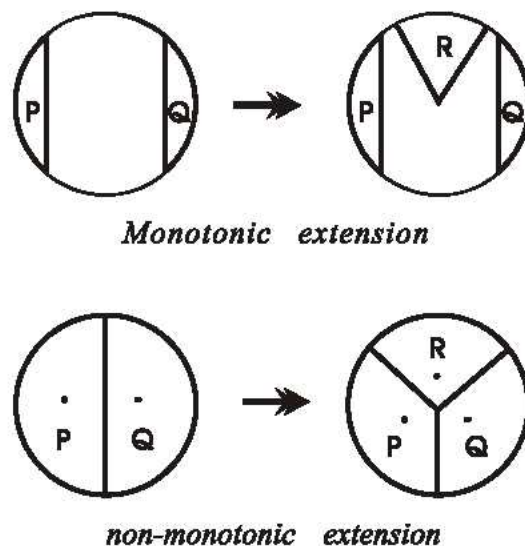


Figure 1: Monotonic and non-monotonic extensions of a (lexicalized) system of concepts

2.4 The persistence of anomaly

Lexical semantics has to account for semantic contradictions such as **married spinster*, **female bachelor*, **reddish green* and for other types of semantic anomalies as exemplified by the famous **Colorless green ideas sleep furiously*. Usually, SEMANTIC ANOMALY of an expression is defined as logical incompatibility of (some part of) the formal translation of the expression taken in union with a given system Γ of definitions and/or meaning postulates (e.g. McCawley 1971). Explicating incompatibility in terms of inconsistency and inconsistency in terms of contradictory entailments makes it possible to derive a second order property which I call the PERSISTENCE OF ANOMALY.

The persistence of anomaly comes in two variants:

- ◆ if we add some new axioms to Γ , then any former anomaly persists
- ◆ if a (propositional) formula is anomalous, then every other formula that implies it is anomalous as well.⁵

Both varieties seem to be satisfied empirically. It would be very surprising if the anomaly of **married bachelor* could be canceled by learning the meaning of several new words. Once an anomaly is established it seems to persist when the system is extended. In a similar sense it would be perplexing if the anomaly of the expression **the idea sleeps* did not persist when the expression is made more specific, e.g. **the new idea sleeps*.

It is straightforward that the notion of semantic anomaly can be converted into a notion of pragmatic anomaly if the system Γ of axioms is assumed to include other sources of knowledge, such as conceptual and ontological knowledge. Not surprisingly, the persistence of anomaly remains in this case.

3. Challenging the standard view

In this section, I will present several phenomena that may raise some doubts about the validity of the four principles just sketched. These phenomena suggest that we take a broader perspective on meaning and include various aspects of utterance interpretation. The examples address the whole spectrum of information shared between lexicon and encyclopedia.

3.1 The principle of compositionality

In section 2.1 we have taken adjectives like *red*, *interesting*, or *straight* as INTERSECTIVE adjectives, and I have illustrated how this fairly simple analysis brings together systematicity and compositionality. Unfortunately, the view that a large range of adjectives behaves intersectively has been shown to be questionable. For example, Quine (1960) notes the contrast between *red apple* (red on the outside) and *pink grapefruit* (pink on the inside), and between the different colors denoted by *red* in *red apple* and *red hair*. In a similar vein, Lahav (1989, 1993) argues that an adjective such as *brown* doesn't make a simple and fixed contribution to any composite expression in which it appears.

In order for a cow to be brown most of its body's surface should be brown, though not its udders, eyes, or internal organs. A brown crystal, on the other hand, needs to be brown both inside and outside. A book is brown if its cover, but not necessarily its inner pages, are mostly brown, while a newspaper is brown only if all its pages are brown. For a potato to be brown it needs to be brown only outside...Furthermore, in order for a cow or a bird to be brown the brown color should be the animal's natural color, since it is regarded as being 'really' brown even if it is painted white all over. A table, on the other hand, is brown even if it is only painted brown and its 'natural' color underneath the paint is, say, yellow. But while a table or a bird are not brown if covered with brown sugar, a cookie is. In short, what is to be brown is different for different types of objects. To be sure, brown objects do have something in common: a salient part that is wholly brownish.

But this hardly suffices for an object to count as brown. A significant component of the applicability condition of the predicate ‘brown’ varies from one linguistic context to another.

(Lahav 1993: 76)

Some authors – for example, Keenan (1974), Partee (1984a), Lahav (1989, 1993) – conclude from facts of this kind that the simplistic view mentioned above must be abolished. As suggested by Montague (1970), Keenan (1974), Kamp (1975) and others, there is a simple solution that addresses such facts in a descriptive way and obeys the principle of compositionality. This solution considers adjectives essentially to be adnominal functors. Such functors, for example, turn the properties expressed by *apple* into those expressed by *red apple*. Of course, such functors have to be defined disjunctively in the manner illustrated in (6):

- (6) RED(X) means roughly the property
- a. of having a red inner volume if X denotes fruits whose inside only is edible
 - b. of having a red surface if X denotes fruits whose outside is edible
 - c. of having a functional part that is red if X denotes tools
- ...

Let us call this the FUNCTIONAL VIEW. It should be stressed that the functional view describes the facts mentioned above only by enumeration. Consequently, it doesn’t account for any kind of systematicity concerning our competence to deal with adjective-noun combinations in an interesting way. Another (notorious) problem with this view has to do with the treatment of predicatively used adjectives. In that case the adjectives must at least implicitly be supplemented by a noun. Various artificial assumptions are necessary which make such a theory inappropriate. We may conclude that compositionality doesn’t necessarily lead to systematicity.

There is a third view about treating the meanings of adjectives, which I call the FREE VARIABLE VIEW. In a certain sense, this view can be seen as preserving the advantages of both the simplistic as well as the functional view, but as overcoming their shortcomings. The free variable view has been developed in considerable detail for the gradable adjectives (see, for example, Bierwisch (1989) and the references given therein). It is well known that the applicability conditions of restricting adjectives that denote gradable properties, such as *tall*, *high*, *long*, *short*, *quick*, *intelligent* vary depending upon the type of object to which they apply. What is high for a chair is not high for a tower and what is clever for a young child is not clever for an adult. Oversimplifying, I can state the free variable view as follows. Similar to the first view, the meanings of adjectives are taken to be one-place predicates. But now we assume that these predicates are complex expressions that contain a free variable. Using an extensional language allowing λ -abstraction, we can represent the adjective *long* (in its contrastive interpretation), for example, as λx LONG(x,X), denoting the class of objects that are long with regard to a comparison class,

which is indicated by the free variable X . At least on the representational level the predicative and the attributive use of adjectives can be treated as in the first view: *The train is long* translates (after λ -conversion) to $\text{LONG}(t, X)$ and *long train* translates to $\lambda x [\text{LONG}(x, X) \wedge T(x)]$. In these formulas t is a term denoting a specific train and T refers to the predicate of being a train.

Free variables are the main instrument for forming underspecified lexical representations. To be sure, free variables simply have the status of place holders for more elaborated subpatterns and expressions containing free variables should be explained as representational schemes. Free variables stand not only as place holders for a comparison class X as just indicated. The view can be generalized to include other types of free variables as well, for example a type of variable connected with the specification of the dimension of evaluation in cases of adjectives such as *good* and *bad* or a type of variable connected with the determination of the object-dependent spatial dimensions in cases of spatial adjectives such as *wide* and *deep*.

Of course, it is not sufficient to postulate underspecified lexical representations and to indicate what the sets of semantically possible specifications of the variables are. In order to grasp natural language interpretation (“conceptual interpretation”), it is also required to provide a proper account of contextual enrichment, explaining how the free variables are instantiated in the appropriate way. Obviously, such a mechanism has to take into consideration various aspects of world and discourse knowledge.

In some particular cases the instantiation of free variables may be done by using ordinary (monotonic) unification. If that works successfully, it may be concluded that the mechanism of contextual enrichment has the feature of compositionality. In other words, the principle of compositionality stated for semantic representations can be transferred to the level of contextually enriched forms. In Blutner (1998), I consider some examples that demonstrate that monotonic unification doesn’t suffice for contextual enrichment.

There are a variety of other examples that demonstrate that our comprehension capacities have salient non-compositional aspects. The most prominent class of examples may be found within the area of SYSTEMATIC POLYSEMY. This term refers to the phenomenon of one lexical unit being associated with a whole range of senses which are related to each other in a systematic way.⁶ The phenomenon has traditionally been thought intractable, and in fact it IS intractable when considered as a problem of lexical semantics in the traditional sense. Related problems with compositionality arise when considering word formation in general (e.g. Aronoff 1976, Bauer 1983) and the interpretation of compounds in particular (e.g. Wu 1990).

3.2 The non-monotonicity of invited inferences

It was the evident divergence between the formal devices $\sim, \wedge, \vee, \supset, (\forall x), (\exists x)$ (in their standard two-valued interpretation) and their natural language counterparts that was the starting point of Grice’s “logic of conversation” (Grice 1967; see Bach and Horn, this volume). In subsequent work these divergences were investigated carefully from an empirical and a theoretical point of view, sometimes adopting Grice’s conceptual framework and sometimes rejecting it. For example, Geis & Zwicky (1971) introduced and discussed the inference scheme they dubbed CONDITIONAL PERFECTION, the notorious

tendency to “perfect” an *if* conditional into the corresponding biconditional (*if and only if*, *iff*). As an example, the utterance of (7a) was claimed to invite the inference of (7b), thus conveying the utterance meaning of (7c).

- (7) a. If you mow the lawn, I’ll give you \$5.
 b. If you don’t mow the lawn, I won’t give you \$5
 c. If and only if you mow the lawn, I’ll give you \$5

In order to account for such inferences, it may be appealing to use rules in the style of natural deduction. For example, we could introduce an inference rule like the following:

- (8) $\frac{\text{if}(P, Q)}{\therefore \text{if}(\sim P, \sim Q)}$

Obviously, (8) can be seen as instantiating the inference from (7a) to (7b). However, in contrast to inference rules like **modus ponens** or **simplification**, the structure-sensitivity of which is never violated (cf. section 2.2), the same doesn’t hold for the schema (8). This was demonstrated by many authors (for a recent survey cf. Horn 2000a). The following examples show situations where the corresponding inferences cannot be drawn:

- (9) a. If John quits, he will be replaced
 b. If John doesn’t quit, he won’t be replaced
 c. If and only if John quits, he will be replaced
- (10) a. If you’re in Toronto, you are in Canada
 b. If you’re not in Toronto, you’re not in Canada
 c. If and only if you’re in Toronto, you are in Canada

There are at least three different strategies of dealing with this observation. The first one is to assume a lexical ambiguity of *if* stipulating two readings: the standard reading and the biconditional reading. The second strategy is to doubt that the tendency of drawing invited inferences in the sense of Geis & Zwicky is a real one, and it aims to reduce the relevant observation exclusively to language-independent factors. The third strategy accepts the reality of these inferences, and, at the same time, acknowledges the non-monotonicity of (parts of) our inferential competence. After pointing out that all three strategies are represented in the literature, Horn (2000a) demonstrates that the third strategy is the most promising one. What’s more, he suggests clarifying the strategy in terms of a non-monotonic operation of pragmatic strengthening⁷ – a suggestion I want to follow in the theoretical part of this chapter (sections 4 and 5.)

Negation in natural language is a rich source of a variety of non-logical inferences (see Horn 1989). Standard examples are SCALAR IMPLICATURES (*Not all of the students came* \approx *Some of them came*). Others are collected under the term NEGATIVE STRENGTHENING.⁸

The latter are concerned with the effect of preferred interpretations that occurs when certain sentence types are negated. In section 5 they are used to explain the basic mechanisms of pragmatic strengthening.

One instance of the phenomenon of negative strengthening arises in connection with gradable adjectives typically occurring as antonyms, such as *{good, bad}*, *{large, small}*, *{happy, unhappy}*. Semantically, the elements of antonym pairs are CONTRARIES, that is, they are mutually inconsistent but do not exhaust the whole spectrum, permitting a non-empty middle ground.

What are the effects of negating gradable adjectives? For the sake of explicitness let's consider the gradable antonyms *happy* and *unhappy*, and assume three possible states of happiness – iconized by 😊, 😞, and 😐. Not unexpectedly, we want to take *happy* as referring to the first state, *unhappy* as referring to the second state, and *neither happy nor unhappy* as referring to the third state.

Let's consider first the effect of negating positive adjectives, starting with a sentence like (11a). Obviously, the preferred interpretation of this sentence is (11c); this corresponds to a logical strengthening of the content of (11a) which is paraphrased in (11b). The discourse (11d) shows that the effect of strengthening (11c) is defeasible. This indicates that the inferential notion that underlies the phenomenon of strengthening ought to be non-monotonic.

- (11) a. I'm not happy
 b. It isn't the case that I'm happy (Entailment) 😞 😊
 c. I'm unhappy (Implicature) 😞
 d. I'm not happy and not unhappy (Defeasibility)

Following Levinson (2000a), the effect of negative strengthening for positive adjectives can be illustrated in the following way:

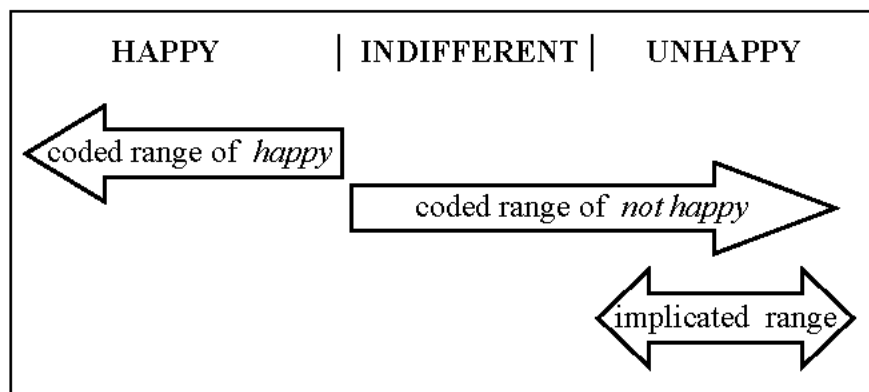


Figure 2: Negative strengthening as implicated contraries

It describes the effect of negative strengthening as implicating contraries from contradictions.

The illustrated shape of negative strengthening is restricted to the positive (unmarked) element of an antonym pair. When considering negative adjectives, deviations from this pattern may be found. The deviations are rather obvious for adjectives with incorporated affixal negation. This leads us to the well-known case of double negation (*Litotes*):

- (12)
- | | |
|--------------------------------------------------------------------------------------------|----------------------|
| a. I'm not unhappy | |
| b. It isn't the case that I'm unhappy | (Entailment) 😊 😐 |
| c. I'm neither happy nor unhappy | (Implicature) 😐 |
| d. I'm rather happy (but not quite as happy as using the expression "happy" would suggest) | (proper Implicature) |
| e. I'm not unhappy, in fact I'm happy | (Defeasibility) |

Admitting only three states on the happiness scale allows only a rather rough approximation of the interpretational effects. The simplest approximation describes negative strengthening as a preference for the middle ground. This is what (12c) expresses. A more appropriate formulation of the effect is given in (12d). For the sake of precision, we had to introduce intermediate states between 😊 and 😐 (on the scale of happiness).

In the following diagram a more adequate illustration of the basic pattern is presented (as described in Horn 1989, Levinson 2000a.)

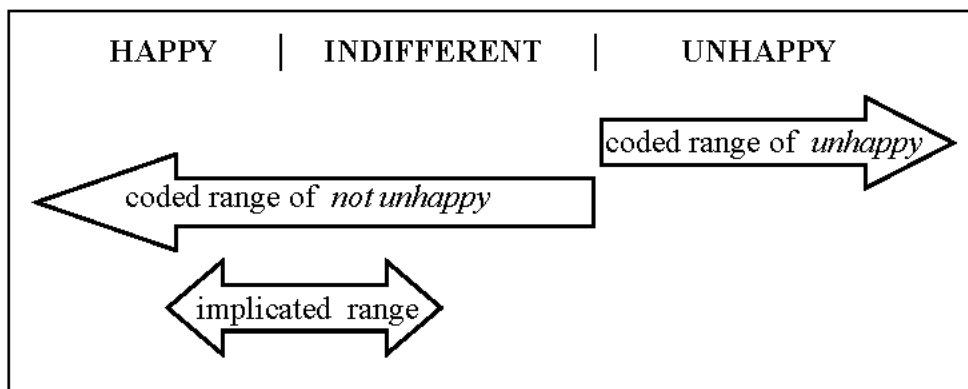


Figure 3: Litotes: when two negatives don't make a positive

As in the case discussed before, the effect of negative strengthening proves defeasible, a fact that requires the underlying inferential notion to be non-monotonic.

The theoretical discussion of the phenomenon of negative strengthening is postponed until section 5, where the inspirational ideas of Horn and Levinson will be outlined and a formal account of their ideas will be given in terms of optimality theory.

3.3 The non-monotonicity of the lexical system

Another general problem that lexical semantics has to address is the phenomenon of LEXICAL BLOCKING. This phenomenon has been demonstrated in a number of examples where the appropriate use of a given expression formed by a relatively productive process is restricted by the existence of a more “lexicalized” alternative to this expression. One case in point was provided by Householder (1971). The adjective *pale* can be combined with a great many color words: *pale green, pale blue, pale yellow*. However, the combination *pale red* is limited in a way that the other combinations are not. For some speakers *pale red* is simply anomalous, and for others it picks up whatever part of the pale domain of red *pink* has not preempted. This suggests that the combinability of *pale* is fully or partially blocked by the lexical alternative *pink*.

Another standard example is the phenomenon of blocking in the context of derivational and inflectional morphological processes. Aronoff (1976) has shown that the existence of a simple lexical item can block the formation of an otherwise expected affixally derived form synonymous with it. In particular, the existence of a simple abstract nominal underlying a given *-ous* adjective blocks its nominalization with *-ity*:

- (13) a. curious – curiosity
 tenacious – tenacity
 b. furious - *furiousity - fury
 fallacious - *fallacity - fallacy

While Aronoff’s formulation of blocking was limited to derivational processes, Kiparsky (1982) notes that blocking may also extend to inflectional processes and he suggests a reformulation of Aronoff’s blocking as a subcase of the ELSEWHERE CONDITION: special rules block general rules in their shared domain. However, Kiparsky cites examples of PARTIAL BLOCKING in order to show that this formulation is too strong. According to Kiparsky, partial blocking corresponds to the phenomenon that the special (less productive) affix occurs in some restricted meaning and the general (more productive) affix picks up the remaining meaning (consider examples like *refrigerant* - *refrigerator*, *informant* - *informer*, *contestant* - *contester*). To handle these and other cases Kiparsky (1983) formulates a general condition which he calls AVOID SYNONYMY: “The output of a lexical rule may not be synonymous with an existing lexical item”.

Working independently of the Aronoff-Kiparsky line, McCawley (1978) collects a number of further examples demonstrating the phenomenon of partial blocking outside the domain of derivational and inflectional processes. For example, he observes that the distribution of productive causatives (in English, Japanese, German, and other languages) is restricted by the existence of a corresponding lexical causative. Whereas lexical causatives (e.g. (14a)) tend to be restricted in their distribution to the stereotypic causative situation (direct, unmediated causation through physical action), productive (periphrastic) causatives tend to pick up more marked situations of mediated, indirect causation. For example, (14b) could be used appropriately when Black Bart caused the sheriff’s gun to backfire by stuffing it with cotton.

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

The phenomenon of blocking can be taken as evidence demonstrating the apparent non-monotonicity of the lexical system. This becomes pretty clear when we take an ontogenetic perspective on the development of the lexical system. Children overgeneralize at some stage while developing their lexical system. For example, they acquire the productive rule of deriving deverbal adjectives with *-able* and apply this rule to produce *washable*, *breakable*, *readable*, but also *seeable* and *hearable*. Only later, after paired forms like *seeable/visible* and *hearable/audible* have coexisted for a while, will the meanings of the specialized items block the regularly derived forms. Examples of this kind suggest that the development of word meanings cannot be described as a process of accumulating more and more denotational knowledge in a monotonic way. Instead, there are highly non-monotonic stages in lexical development. At the moment, it is not clear whether this ontogenetic feature must be reflected in the logical structure of the mental lexicon. Rather, it is possible that pragmatic factors (such as Gricean rules of conversation) play an important role in determining which possible words are actual and what they really denote (McCawley 1978, Horn 1984a, Dowty 1979).

3.4 The non-persistence of (pragmatic) anomaly

Take the well-known phenomenon of CONCEPTUAL GRINDING, whereby ordinary count nouns acquire a mass noun reading denoting the stuff the individual objects are made of, as in *Fish is on the table* or *Dog is all over the street*. There are several factors that determine whether grinding may apply, and, more specifically, what kind of grinding (meat grinding, fur grinding, universe grinding, ...) may apply. Some of these factors have to do with the conceptual system, while others are language-dependent (cf. Nunberg & Zaenen 1992; Copestake & Briscoe 1995).

One of the language-dependent factors affecting the grinding mechanism is lexical blocking. For example, in English the specialized mass terms *pork*, *beef*, *wood* usually block the grinding mechanism in connection with the count nouns *pig*, *cow*, *tree*. This explains the contrasts given in (15).

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

The important point is the observation that blocking is not absolute but may be canceled under special contextual conditions. That is, we find cases of DEBLOCKING. Nunberg & Zaenen (1992) consider the following example:

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

They argue that “what makes *beef* odd here is that the interdiction concerns the status of the animal as a whole, and not simply its meat. That is, Hindus are forbidden to eat beef only because it is cow-stuff.” (Nunberg & Zaenen 1992: 391). Examples of this kind strongly suggest that the blocking phenomenon is pragmatic in nature. Furthermore, these examples suggest that (pragmatic) anomaly does not necessarily persist when specific contextual information is added. Copestake & Briscoe (1995) provide further examples that substantiate this claim.

4. Lexical underspecification and pragmatic strengthening

In the late 70’s a renewed interest in the formal treatment of indexical expressions (*I, you, he, here, now, that, that book*, etc.) within model-theoretic semantics can be observed, inspired primarily by the work of Montague (e.g. Montague 1970). The basic idea was to fill in some of the gaps in earlier work by Carnap and others by introducing aspects of context into formal semantics. As a result of these efforts, something like a classical theory of context-dependency originated.⁹

Based on this work, researchers in lexical semantics stressed the importance of the distinction between lexicon and encyclopedia. Going beyond the aspects of meaning typically investigated by a contrastive analysis of lexemes within the standard view, they proposed different analyses dealing with the meaning of words within concrete conceptual and contextual settings. Examples *par excellence* are Nunberg’s (1979) and Sag’s (1981) analyses of predicate transfer, the work of Bartsch (1987) on systematic polysemy, Bierwisch’s (1983, 1989) two-level semantics¹⁰, and Pustejovsky’s (1991, 1993, 1995) theory of the generative lexicon. Most of these rather diverse accounts can be seen as different ways of closing the gap between lexical semantics and pragmatics.

4.1 The coercion view and the view of radical underspecification

The most notable feature in drawing the line between semantics and pragmatics is the principle of compositionality. Many researchers agree in seeing compositionality as a principle satisfied at the level of semantic representation and violated at the level of utterance interpretation. Respecting the non-compositionality of utterance interpretation, several of these researchers seem to consider it virtuous and advantageous to deviate from compositionality in a minimalist way. A typical approach following this path of virtue is the so-called COERCION VIEW (cf. Pustejovsky 1993, 1995), which is summarized here:

(17) The Coercion View

- a. Every lexical unit determines a primary conceptual variant that can be grasped as its (literal) meaning.
- b. The combinatorial system of language determines how the lexical units are combined into larger units (phrases, sentences.)
- c. There is a system of type and sortal restrictions that determines whether the

resulting structures are well-formed.

- d. There is a generative device (called TYPE/SORT COERCION) that tries to overcome type or sortal conflicts that may arise by strict application of the combinatorial system of language. The coercion device is triggered (only) by type or sort violations.

In Blutner (1998, 2002) several problems about the coercion view in general and Pustejovsky's account in particular are discussed. Taken together, several key issues – the combinatorial explosion puzzle, the restrictiveness of the coercion mechanism, the problem of defeasibility, an inflation of shifting operations – suggest that it might be promising to look for an alternative view.

The most promising alternative to the coercion view is a conception that is called the RADICAL UNDERSPECIFICATION VIEW. Contrasting sharply with the coercion view, this position is more radically founded on underspecified representation, and it explicitly makes use of a pragmatic mechanism of contextual strengthening.

(18) The Radical Underspecification View

- a. Every lexical unit determines an underspecified representation (i.e. a representation that may contain, for example, place holders and restrictions for individual and relational concepts)
- b. The combinatorial system of language determines how lexical units are combined into larger units (phrases, sentences)
- c. There is a system of type and sortal restrictions that determines whether structures of a certain degree of (under)specification are well-formed
- d. There is a mechanism of contextual enrichment (pragmatic strengthening based on contextual and encyclopedic knowledge). This inferential mechanism is controlled by cost factors and need not be triggered by type or sort violations

The view of radical underspecification shares several ideas with alternative accounts: it takes the distinctions between lexicon and encyclopedia, and likewise between semantics and pragmatics, as very important. Moreover, it takes the features of compositionality, monotonicity, and persistence of anomaly as crucial characteristics marking out the domain of semantics. However, in contrast to Pustejovsky's generative lexicon and Bierwisch's two-level semantics, the view of radical underspecification disregards monotonic unification and type/sort coercion as mechanisms of contextual enrichment. Instead, it explores alternative proposals that stress open-ended default inference on real world knowledge. Here is a small collection of candidates that may provide a suitable mechanism for the contextual enrichment of underspecified representations:

- ◆ Defaults as rules for filling in information gaps (see various papers in van Deemter & Peters (eds.) 1996)
- ◆ Discourse interpretation based on a default conditional logic (e.g. Lascarides &

Asher 1993)

- ◆ Persistent Default-Unification (Lascarides et al. 1995)
- ◆ Weighted abduction (Hobbs et al. 1993; cf. Hobbs, this volume)
- ◆ Conversational implicature and lexical pragmatics (Blutner, Leßmöllmann & van der Sandt 1996, Blutner 1998)

In the next section I will refer to the last-mentioned account only, and I will outline how this account may solve some of the problems stated above.

4.2 Conversational implicature and lexical pragmatics

For Griceans, conversational implicatures are those non-truth-functional aspects of utterance interpretation which are conveyed by virtue of the assumption that the speaker and the hearer are obeying the COOPERATIVE PRINCIPLE of conversation, and, more specifically, various CONVERSATIONAL MAXIMS of quantity, quality, relation and manner. While the notion of conversational implicature doesn't seem hard to grasp intuitively, it has proven difficult to define precisely. An important step in reducing and explicating the Gricean framework has been made by Atlas & Levinson (1981) and Horn (1984a). Taking Quantity as starting point they distinguish between two principles, the Q-principle and the I-principle (termed the R-principle by Horn 1984a). Simple informal formulations of these principles are as follows:

Q-principle:

- ◆ Say as much as you can (given I) (Horn 1984a: 13)
- ◆ Make your contribution as informative (strong) as possible (Matsumoto 1995: 23)
- ◆ Do not provide a statement that is informationally weaker than your knowledge of the world allows, unless providing a stronger statement would contravene the I-principle (Levinson 1987b: 401)

I-principle:

- ◆ Don't say more than you must (given Q) (Horn 1984a: 13)
- ◆ Say as little as necessary, i.e. produce the minimal linguistic information sufficient to achieve your communicational ends (bearing the Q-principle in mind) (Levinson 1987b: 402)
- ◆ Read as much into an utterance as is consistent with what you know about the world (Levinson 1983: 146-47)

Obviously, the Q-principle corresponds to the first part of Grice's quantity maxim (*Make your contribution as informative as required*), while it is argued (cf. Horn 1989, 1993) that the countervailing I-principle collects the second part of the quantity maxim (*Do not make your contribution more informative than is required*), the maxim of relation and at least two of the manner submaxims, "Be brief" and "Be orderly". As Horn (1984a) seeks to demonstrate, the two principles can be seen as representing two competing forces, one force of *unification* minimizing the Speaker's effort (I-principle), and one force of *diversification* minimizing the Auditor's effort (Q-principle).

Conversational implicatures which are derivable essentially by appeal to the Q-principle are called Q-based implicatures. Standard examples are scalar implicatures and clausal implicatures. I-based implicatures, derivable essentially by appeal to the I-principle, can be

generally characterized as enriching what is said via inference to a rich, stereotypical interpretation (cf. Gazdar 1979, Atlas & Levinson 1981, Horn 1984a, Levinson 2000a)

In my opinion, the proper treatment of conversational implicature crucially depends on the proper formulation of the Q- and the I-principle. The present explication (cf. Blutner 1998) rests on the assumption that the semantic description of an utterance is an underspecified representation f determining a wide range of possible enrichments m , one of which covers the intended content. There are different possibilities to make explicit what possible enrichments are: the idea of abductive specification may be useful (e.g. Hobbs et al. 1993), and likewise the idea of non-monotonic unification (e.g. Lascarides et al. 1995). Both mechanisms make use of the notion of *common ground*, which informally can be introduced as an information state containing all the propositions shared by several participants, including general world and discourse knowledge. I will not be very specific about the device that generates possible enrichments. For the sake of convenience, I simply assume a function **Gen** that determines for each common ground σ what the possible enrichments of f are. In other words,

- (19) $\langle f, m \rangle$ is called a **possible enrichment pair** (or *pep*) iff $\langle f, m \rangle \in \mathbf{Gen}_\sigma$, i.e. m can be generated from f by means of a common ground σ .

The other important component that is necessary to reconstruct the essence of conversational implicature is the evaluation component. It evaluates *peps* and typically is defined by a cost function $\underline{c}(f, m)$ (cf. Blutner 1998). For example, in weighted abduction (Hobbs et al. 1993) this function reflects the *proof* cost for deriving an interpretation m from the underspecified form f .¹¹ For the present aims it is not necessary to have the numerical values of this cost function. What is sufficient is an ORDERING RELATION \succ (*being more harmonic, being more economical*) defined on the *peps*.¹² To be sure, the concrete realization of this ordering relation relates to a variety of different graded factors such as informativity, relevance, and effort, and is a matter for empirical investigation (cf. Ducrot 1972, Merin 1999, van Rooy to appear a,b.)

In Blutner (1998) it is pointed out that the effect of the Gricean maxims is simply to constrain the relation defined by **Gen** in a particular way. In short, the Q- and the I-principle can be seen as conditions constraining possible enrichment pairs $\langle f, m \rangle$. The precise formulation assumes the availability of the (partial) ordering \succ and formulates a two-way optimization procedure:

- (20) a. $\langle f, m \rangle$ satisfies the Q-principle iff $\langle f, m \rangle \in \mathbf{Gen}_\sigma$ and there is no other pair $\langle f, m' \rangle$ such that $\langle f, m' \rangle \succ \langle f, m \rangle$
- b. $\langle f, m \rangle$ satisfies the I-principle iff $\langle f, m \rangle \in \mathbf{Gen}_\sigma$ and there is no other pair $\langle f, m' \rangle$ such that $\langle f, m' \rangle \succ \langle f, m \rangle$ ¹³

In this (rather symmetrical) formulation, the Q- and the I-principle constrain the *peps* in

two different ways. The I-principle constrains them by selecting the most economic/harmonic enrichments, and the Q-principle constrains them by blocking those enrichments which can be grasped more economically/harmonically by an alternative linguistic input f . Obviously, it is the Q-principle that carries the main burden in explaining the blocking effects discussed in section 3.2.

The important definitions of pragmatic anomaly and conversational implicature can be stated as follows, making use of an auxiliary notion called PRAGMATIC LICENSING:

- (21) a. A *pep* $\langle f, m \rangle$ is called **pragmatically licensed** (in a common ground σ) iff $\langle f, m \rangle$ satisfies the Q- and the I-principle and m is consistent with σ .
- b. An utterance that corresponds to the (underspecified) semantic form f is called **pragmatically anomalous** (in σ) iff there is no pragmatically licensed *pep* $\langle f, m \rangle$.
- c. A proposition p is called a **conversational implicature** of f (in σ) iff p is a classical consequence of $\sigma \cup m$ for each m of a pragmatically licensed *pep* $\langle f, m \rangle$.

It is not difficult to see how the general mechanism of conversational implicature introduced in (21) reflects the four features/phenomena repeated here for convenience:

- ◆ the non-compositional aspect of utterance interpretation
- ◆ the non-monotonicity of conversational implicature
- ◆ the phenomena of blocking and deblocking
- ◆ the general fact that pragmatic anomalies usually don't persist

First, let's consider compositionality. Almost everything in the formulation of conversational implicature has a non-compositional character: the formulation of both the Q-principle and the I-principle is *holistic* in addressing a wide range of alternative expressions; the conceptions of informativeness, surprise (measured in terms of conditional probability), and linguistic complexity are non-combinatorial and cannot be reduced to the corresponding properties of the parts of an expression (cf. Blutner 1998).

Next, our system deals with non-monotonicity by basing the notion of conversational implicature on preferred interpretations (via the optimization of *peps*). It is the old insight of McCarthy (1980), Shoham (1988), and others that the idea of preferred interpretations establishes a non-monotonic (cumulative) inferential relation.

Third, our system deals with blocking and deblocking. The crucial mechanism involved is due to the Q-principle. In the same way, the present system captures the field-effects, which are very important if the extensions of lexical concepts are considered.

The fourth and last point concerns the persistence of anomalies. The general definition of pragmatic anomaly doesn't simply define this notion as some kind of inconsistency. Instead, non-representational parameters (such as surprise, cue validity, relevance, frequency of use, etc.) are crucially involved in controlling the selection and suppression of possible enrichments. Within this setting, typically some kind of garden path effect may arise. This constitutes pragmatic anomaly (for details, see Blutner 1998.)

5. Optimality theory and lexical pragmatics

The situated meanings of many words and simple phrases are combinations of their lexical meanings proper and some superimposed conversational implicatures. In the previous sections we have suggested to represent lexical meanings by means of underspecified forms and to take compositionality, monotonicity and the persistence of anomaly as bounding the domain of semantics proper. On the other hand, a mechanism of pragmatic strengthening was suggested which crucially makes use of non-representational parameters that are described by a certain ordering relation. It is the use of an optimization procedure that gives pragmatics its holistic flourish, systematically destroys the listed features, and relegates them to the level of semantics proper.

Before we come to the treatment of examples, we will show the close relationship between the formulation of pragmatic strengthening given in the previous section and recent developments in OPTIMALITY THEORY (OT)

5.1 Bidirectional OT and pragmatic strengthening

OT is a linguistic framework that is not only of interest to phonologists but has likewise attracted students of morphology, syntax and natural language interpretation. As pointed out by Anttila & Fong (2000), current work in optimality theoretic syntax and semantics has been concerned with two closely related questions.

- ◆ OT Syntax: Given a semantic input, what is its optimal expression?
- ◆ OT Semantics: Given a syntactic input, what is its optimal interpretation?

OT syntax takes the point of view of the speaker (the EXPRESSIVE perspective): given a semantic input, the goal is to select the optimal syntactic expression for this input among a well-defined set of candidate expressions (see e.g. Grimshaw 1997, Bresnan 2001). OT semantics takes the point of view of the hearer (the INTERPRETIVE perspective): given a syntactic input, the goal is to select the optimal semantic interpretation among a set of candidate interpretations (see e.g. de Hoop & de Swart 1998, Hendriks & de Hoop (to appear), de Hoop 2000). In Blutner (1999) I argue that this design of OT – taking the different perspectives isolated from each other – is inappropriate and too weak in a number of cases. What I proposed is bidirectional optimization where both types of optimization are carried out simultaneously.

The formulation of the Q- and I-principle in (20) makes it quite clear that a bidirectional optimality framework – integrating expressive (Q) and interpretive (I) optimization – is an appropriate tool to reconstruct the Gricean mechanism of pragmatic strengthening.¹⁴ Consider the following conception of OPTIMAL form-meaning pairs which is a straightforward reformulation of (20). It is dubbed the STRONG version of bidirectional OT.

(22) **Bidirectional OT (Strong Version)**

A form-meaning pair $\langle f, m \rangle$ is called **optimal** iff $\langle f, m \rangle \in \text{Gen}_\sigma$ and

- (Q) there is no other pair $\langle f', m \rangle$ such that $\langle f', m \rangle \succ \langle f, m \rangle$
 (I) there is no other pair $\langle f, m' \rangle$ such that $\langle f, m' \rangle \succ \langle f, m \rangle$

The crucial notion of PRAGMATICALLY LICENSED *peps* (22a) now transforms into

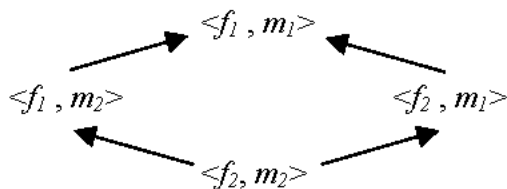
- (22) a'. A *pep* $\langle f, m \rangle$ is called **pragmatically licensed** (in a common ground σ) iff $\langle f, m \rangle$ is optimal and m is consistent with σ .

At first glance, using the bidirectional competition technique can be seen as just establishing the very same ideas presented in Blutner (1998) by means of a more broadly acknowledged and more well-known basis. However, that is not the whole story. We have to acknowledge that the framework of OT gives us a much wider perspective for relating natural language comprehension, language acquisition (Tesar & Smolensky 2000), and language change (e.g. Haspelmath 1999b). Furthermore, there are important explorations concerning the concrete realization of the (harmonic) ordering relation (e.g. Beaver 2000, Zeevat 1999a,b, Aissen 2000). Taking the broader perspective and the more rigorous formalization, the use of OT may give the enterprise of Radical Pragmatics in general and Lexical Pragmatics in particular a new impulse.

In standard OT the ordering relation between elements of the generator is established via a system of ranked constraints. These constraints are typically assumed to be output constraints, i.e. they may be either satisfied or violated by an output form. In the bidirectional framework changing perspectives are possible. This means that an output under one perspective can be seen as an input under the other perspective. Therefore, it is plausible to assume output AND input constraints. Seeing the input as a linguistic form that conveys phonological, syntactic and semantic information, constraints on inputs are typically markedness conditions evaluating the HARMONY of forms. On the other hand, the output (i.e. the result of contextual enriching) is evaluated by constraints that determine its coherence and informativeness (with regard to a context σ).

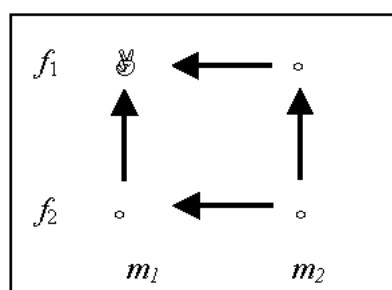
I will now give a very schematic example in order to illustrate some characteristics of bidirectional OT. Assume that we have two forms f_1 and f_2 which are semantically equivalent. This means that **Gen** associates the same meanings with them, say m_1 and m_2 . We stipulate that the form f_1 is less complex (marked) than the form f_2 and that the interpretation m_1 is less complex (marked) than the interpretation m_2 . From these differences of markedness with regard to the levels of syntactic forms/semantic interpretations, the following ordering relation between form-meaning pairs can be derived:

- (23) a. $\langle f_1, m_1 \rangle \succ \langle f_2, m_1 \rangle$
 b. $\langle f_1, m_2 \rangle \succ \langle f_2, m_2 \rangle$
 c. $\langle f_1, m_1 \rangle \succ \langle f_1, m_2 \rangle$
 d. $\langle f_2, m_1 \rangle \succ \langle f_2, m_2 \rangle$



Using Dekker's & van Rooy's (1999) notation, a bidirectional OT diagram can be construed, nicely representing the preferences between the pairs (the arrows point to the preferred pair.) More importantly, such diagrams give an intuitive visualization for the optimal pairs of (strong) bidirectional OT: they are simply the hollows if we follow the arcs.¹⁵ The optimal pairs are marked with the symbol \mathfrak{B} in the diagram.

(24)



The scenario just installed describes the case of **total** blocking where some forms (e.g. **furiousness*, **fallaciousness*) do not exist because others do (*fury*, *fallacy*). However, as noted in section 3.3, blocking is not always total but may be **partial**, in that only those interpretations of a form are ruled out that are pre-empted by a “cheaper” competing form.

Cases of total and partial blocking are not only found in morphology, but in syntax and semantics as well (cf. Atlas & Levinson 1981, Horn 1984a, Williams 1997). The general tendency of partial blocking seems to be that “unmarked forms tend to be used for unmarked situations and marked forms for marked situations” (Horn 1984a: 26) – a tendency that Horn terms the DIVISION OF PRAGMATIC LABOR (see Horn, this volume).

We have seen that the strong form of bidirectionality describes total blocking and doesn't account for partial blocking. There are two principal possibilities for avoiding the fatal consequences of total blocking. The first possibility is to make some stipulations concerning **Gen** in order to exclude equivalent semantic forms. The second possibility is to weaken the notion of (strong) optimality in a way that allows us to derive Horn's division of pragmatic labor in a principled way by means of a sophisticated optimization procedure.

In Blutner (1998,1999) I argue that the second option is much more practicable and theoretically interesting. I proposed a recursive variant of bidirectional optimization (also called WEAK bidirection) which was subsequently simplified by Jäger (2000). Here is Jäger's formulation:

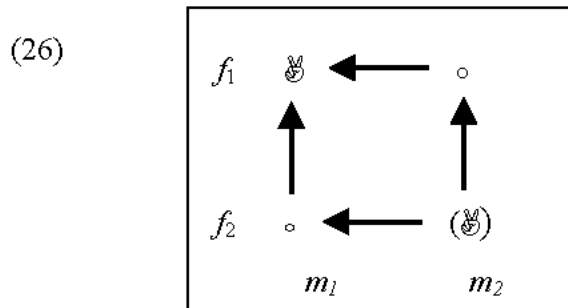
(25) Bidirectional OT (Weak Version)

A form-meaning pair $\langle f, m \rangle$ is called **super-optimal** iff $\langle f, m \rangle \in \mathbf{Gen}_\sigma$ and

(Q) there is no other super-optimal pair $\langle f', m' \rangle : \langle f', m' \rangle \succ \langle f, m \rangle$

- (I) there is no other super-optimal pair $\langle f, m' \rangle : \langle f, m' \rangle \succ \langle f, m \rangle$

Under the assumption that \succ is transitive and well-founded, Jäger (2000) proved that (25) is a sound recursive definition and he showed its equivalence with the formulation in Blutner (1998, 1999). In addition, he proved that each pair that is **optimal** (strong bidirection) is **super-optimal** (weak bidirection) as well, but not vice versa. Hence, weak bidirection gives us a chance to find additional super-optimal solutions. For example, weak bidirection allows marked expressions to have an optimal interpretation, although both the expression and the situations they describe have a more efficient counterpart. To make this point clear, consider again the situation illustrated in (24), but now applying the weak version of bidirectional optimization as demonstrated in the diagram (27). In order to make things more concrete we can take f_1 to be the lexical causative form (14a), f_2 the periphrastic form (14b), m_1 direct (stereotypic) causation and m_2 indirect causation.



The solution $\langle f_1, m_1 \rangle$ comes out as the strong solution and can be taken as the initial point of recursion. The pairs $\langle f_1, m_2 \rangle$ and $\langle f_2, m_1 \rangle$, respectively, are both blocked by this solution.

We have seen that the **strong** version cannot explain why the marked form f_2 has an interpretation. The **weak** version, however, can explain this fact. Moreover, it explains that the marked form f_2 gets the atypical interpretation m_2 . This comes out since the pair $\langle f_2, m_2 \rangle$ is **super-optimal** according to definition (26): It is never blocked by another super-optimal pair.¹⁶ In this way, the weak version of bidirection accounts for Horn's "division of pragmatic labor". Jäger (2000) has shown that this pattern can be generalized to systems where more than two forms are associated by **Gen** with more than two interpretations. In the general case, we start by determining the optimal pairs. Then we drop the rows and columns corresponding to the optimal pair(s) and apply the same procedure to the reduced tableau.¹⁷

5.2 Negative strengthening

In section 3.2 a concise description of the phenomenon of negative strengthening was given. Now I will bring this phenomenon into play in order to illustrate the general mechanism of pragmatic strengthening, which is formulated by using the method of bidirectional optimization.

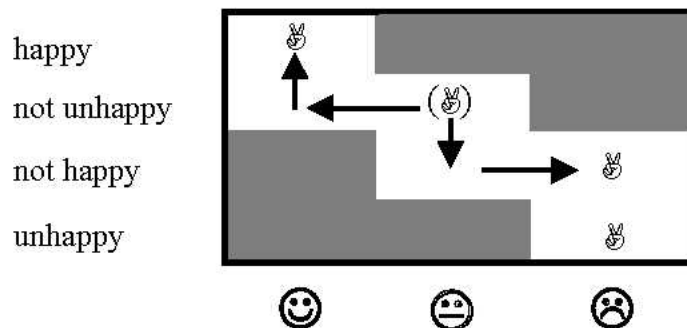
In the analysis of Horn (1989) and Levinson (2000a) there are some types of negative

strengthening that are obviously attributable to the I/R-principle. A clear case is the negation of positive adjectives, which was described in connection with example (11). Here the I/R principle leads to a pragmatic strengthening effect excluding the middle ground and implying the contrary.

The situation is not so clear in the case of adjectives with incorporated affixal negation such as in example (12). Whereas Horn (1984a, 1989) attributes the observed effect of negative strengthening to the interaction between Q and R, Levinson stipulates a third pragmatic principle, the M(anner)-principle: “what’s said in a abnormal way, isn’t normal; or marked message indicates marked situation.” (Levinson 2000a: 33). Obviously, this principle expresses the second half of Horn’s division of pragmatic labor (see also Horn 1991a on double negation and the division of pragmatic labor).¹⁸

Let’s see now how bidirectional OT accounts for the effects of negative strengthening. The bidirectional tableau (28) shows the competing candidate forms in the left column. (Take the candidate entries as shortcuts for complete sentences; for example take *happy* as abbreviating *I’m happy*, etc.). The other columns are for the three possible states of happiness considered in this simplified analysis. The gray areas in the tableau indicate which form-interpretation pairs are excluded by the compositional mode of truth-functional semantics. For example, *I’m not unhappy* is assumed to exclude the state iconized by ☹.

(27)



The preferences between the form-interpretation pairs are due to markedness constraints for forms and markedness constraints for interpretations, respectively.

With regard to the forms, we simply assume that the number of negation morphemes is the crucial indicator. The corresponding preferences are indicated by the vertical arrows. (Note that *not happy* and *unhappy* aren’t differentiated in terms of markedness – a rough simplification, of course.)

With regard to the states, we assume that they are decreasing in markedness towards both ends of the scale, assigning maximal markedness to the middle ground. Although this assumption seems not implausible from a psycholinguistic perspective, I cannot provide independent evidence for it at the moment. In (27), the corresponding preferences are indicated by the horizontal arrows.

Now it is a simple exercise to find out the optimal solutions – indicated by ☺. One

optimal solution pairs the sentence *I'm not happy* with the interpretation \ominus . This solution corresponds to the effect of negative strengthening that is attributable to the I/R-principle. The other two optimal solutions reflect the truth conditions of *I'm happy/unhappy*.

Most interesting, there is an additional super-optimal solution, indicated by \otimes . It pairs the sentence *I'm not unhappy* with the interpretation \ominus . This corresponds to the effect of negative strengthening in the case of LITOTES, normally attributed to Levinson's (2000a) M-principle or Horn's division of pragmatic labor (cf. Horn 1991a). As already stressed, this solution comes out as a natural consequence of the weak form of bidirection, which can be seen as a formal way of describing the interactions between Q and I/R.

It's an interesting exercise to introduce more than three states of happiness and to verify that the proper shape of implicature as indicated in Figure 3 can be approximated. More importantly, in the context of litotes it seems necessary to account for the effect of gradient acceptability and continuous scales. Using a stochastic evaluation procedure, Boersma (1998) and Boersma & Hayes (2001) did pioneering work in this field, which should be exploited in the present case.

The other prominent class of examples that exhibit the effect of negative strengthening concerns the phenomenon of neg-raising, i.e. the tendency for negative main sentences with subordinate clauses to be read as negations of the subordinate clause (cf. Horn 1989, Levinson 2000a). It seems fruitful to analyze the phenomenon using the same technique as described above.

6. Conclusions

Investigating the interactions between the (mental) lexicon and pragmatics, we have pointed out that situated meanings of many words and simple phrases are combinations of their lexical meanings proper and some superimposed conversational implicatures. In particular, we have suggested representing lexical meanings by means of underspecified forms and using compositionality, monotonicity and the persistence of anomaly for demarcating the borderline between semantics proper and pragmatics. A mechanism of pragmatic strengthening was suggested which crucially makes use of "non-representational" parameters that are described by preferential relations, such as information scales or salience orderings. The basic pragmatic mechanism can be expressed within the framework of bidirectional OT.

This approach may provide a principled account of several lexical-pragmatic phenomena that are currently being investigated: negative strengthening for graded antonym pairs (Horn 1989, Levinson 2000a), the effects of neg-raising (Horn 1989), the pragmatics of adjectives (Lahav 1993, Blutner 1998), the distribution of lexical and productive causatives (McCawley 1978, Horn 1984a), the pragmatics of dimensional designation (Blutner & Solstad 2000), the interpretation of compounds (Meyer 1993) and many phenomena presently discussed within the framework of Cognitive Semantics.

The main advantage of bidirectional OT is that it helps us to put in concrete terms what the **requisites** are for explaining the peculiarities of negative strengthening, neg-raising and the other phenomena under discussion. What are the relevant cognitive scales? How do we measure morpho-syntactic markedness? How do we measure the values of

probabilistic parameters that control and organize conceptual knowledge (salience, cue validity)? The latter proves essential when it comes to consider word formation and the investigation of different types of polysemy.

An important challenge for the present view is the work done in relevance theory (e.g. Sperber & Wilson 1986a, Carston 1998b, 1999, this volume). Although I prefer a variant of Atlas's, Levinson's and Horn's frameworks, that doesn't mean that I am taking a stand against relevance theory. Rather, it seems desirable and possible to integrate the major insights from relevance theory into the present view. As a kind of meta-framework, optimality theory can help to realize this integrative endeavor and to bring the two camps closer to each other. Recently, van Rooy (to appear b, c) has taken the first important steps in this direction.

REFERENCES

- Aissen, Judith (2000), 'Differential object marking: iconicity vs'. Manuscript UCSC. Stanford, CA.
- Anttila, Arto & Vivienne Fong (2000), 'The partitive constraint in optimality theory'. Paper available from the Rutgers Optimality Archive. ROA 416-09100.
- Atlas, Jay David & Stephen C. Levinson (1981), 'It-clefts, informativeness and logical form'. In Peter Cole (ed.), *Radical Pragmatics*. Academic Press. New York. 1-61.
- Aronoff, Mark. (1976), *Word Formation in Generative Grammar*. MIT Press. Cambridge, Mass.
- Bartsch, Renate (1987), 'Context-dependent interpretations of lexical items'. In Renate Bartsch, Johan van Benthem, and Peter van Emde-Boas (eds.), *Semantics and Contextual Expressions*. Foris. Dordrecht.
- Bauer, Laurie (1983), *English Word-Formation*. Cambridge University Press. Cambridge.
- Beaver, David (2000), 'Centering and the optimization of discourse'. Manuscript Stanford University. Stanford, CA.
- Bierwisch, Manfred (1983), 'Semantische Einheiten und konzeptuelle Repräsentation lexikalischer Einheiten'. In *Untersuchungen zur Semantik*. Akademie-Verlag. Berlin, 61-99.
- Bierwisch, Manfred (1989), 'The semantics of gradation'. In Manfred Bierwisch & Ewald Lang (eds.), *Dimensional Adjectives*. Springer-Verlag. Berlin. 71-261.
- Blutner, Reinhard (1998), 'Lexical Pragmatics'. *Journal of Semantics* 15: 115-162.
- Blutner, Reinhard (1999), 'Some aspects of optimality in natural language interpretation'. In Helen de Hoop and Henriëtte de Swart (eds.), *Papers on Optimality Theoretic Semantics*. Utrecht University of Linguistics. Utrecht. 1-21.
- Blutner, Reinhard (2002), 'Lexical semantics and pragmatics'. *Linguistische Berichte* 10, 27-58.

- Blutner, Reinhard, Annette Leßmöllmann, & Rob van der Sandt (1996): 'Conversational implicature and lexical pragmatics'. In *Proceedings of the AAAI Spring Symposium on Conversational Implicature*. Stanford. 1-9.
- Blutner, Reinhard & Torgrim Solstad (2000), 'Dimensional designation: a case study in lexical pragmatics'. In Reinhard Blutner and Gerhard Jäger (eds.), *Studies in Optimality Theory*. Linguistics in Potsdam. Potsdam. 30-40.
- Boersma, Paul (1998), *Functional phonology*. Holland Academic Graphics. The Hague.
- Boersma, Paul & Bruce Hayes (2001), 'Empirical Tests of the gradual learning algorithm'. *Linguistic Inquiry* 32: 45-86.
- Bresnan, Joan (2001), 'The emergence of the unmarked pronoun'. In BLS 23. Berkeley, CA. Berkeley Linguistic Society. Also: ROA-179-0297.
- Carston, Robyn (1999), 'The semantics/pragmatics distinction: a view from relevance theory'. *UCL Working Papers in Linguistics* 10: 1-30.
- Carston, Robyn (1998b). 'Informativeness, relevances and scalar implicature'. Manuscript. University College London.
- Copstake, Ann & Briscoe, Ted (1995), 'Semi-productive polysemy and sense extension'. *Journal of Semantics* 12: 15-67.
- Cruse, D.A. (1986), *Lexical Semantics*. Cambridge University Press. Cambridge.
- Deane, Paul D. (1988): 'Polysemy and cognition'. *Lingua* 75: 325-361.
- de Hoop, Helen (2000), 'Optimal scrambling and interpretation'. In H. Bennis, M. Everaert, and E. Reuland (eds.), *Interface Strategies*. KNAW. Amsterdam. 153-168.
- de Hoop, Helen & Henriëtte de Swart (1998), 'Temporal adjunct clauses in optimality theory'. Manuscript, OTS Utrecht.
- Dekker, Paul & van Rooy, Robert (1999), 'Optimality theory and game theory: some parallels'. In Helen de Hoop and Henriëtte de Swart (eds.), *Papers on Optimality Theoretic Semantics*. Utrecht University of Linguistics. Utrecht. 22-45.
- Dowty, David (1979), *Word Meaning and Montague Grammar*. Kluwer. Dordrecht.
- Ducrot, Oswald (1972), *Dire et ne pas dire*. Hermann. Paris.
- Fodor, Janet D. (1977), *Semantics: Theories of Meaning in Generative Grammar*. Crowell. New York.
- Fodor, Jerry A. & Zenon W. Pylyshyn (1988), 'Connectionism and cognitive architecture: a critical analysis'. *Cognition* 28: 3-71.
- Gärdenfors, Peter (1993), 'The emergence of meanings'. *Linguistics and Philosophy* 16: 285-309.
- Gärdenfors, Peter (2000), *Conceptual Spaces: The Geometry of Thought*. The MIT Press. Cambridge, Mass.
- Gazdar, Gerald (1979), *Pragmatics*. Academic Press. New York.
- Geis, Michael L. & Arnold M. Zwicky (1971), 'On invited inference'. *Linguistic Inquiry* 2: 561-566.
- Grice, H. Paul (1968): *Logic and Conversation*. Text of Grice's William James lectures at Harvard University. Published in: Paul Grice (1989): *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.
- Grimshaw, Jane (1997), 'Projection, heads, and optimality'. *Linguistic Inquiry* 28: 373-422.

- Haspelmath, Martin (1999b), 'Optimality and diachronic adaptation'. *Zeitschrift für Sprachwissenschaft* **18**: 180-205.
- Hendriks, Petra & Helen de Hoop (to appear), 'Optimality theoretic semantics'. *Linguistics and Philosophy*.
- Hobbs, Jerry, Mark E. Stickel, Douglass E. Appelt, & Paul Martin (1993), 'Interpretation as abduction'. *Artificial Intelligence* **63**: 69-142.
- Horn, Laurence R. (1984a), 'Toward a new taxonomy for pragmatic inference: Q-based and R-based implicatures'. In D. Schiffrin (ed), *Meaning, Form, and Use in Context*. Georgetown University Press. Washington. 11-42.
- Horn, Laurence R. (1989), *A natural history of negation*. University of Chicago Press. Chicago.
- Horn, Laurence R. (1991a), 'Duplex negatio affirmat...: The economy of double negation'. In *Papers from the Parasession on Negation*, CLS 27/2, 78-106. Chicago: CLS.
- Horn, Laurence R. (2000), 'From if to iff: conditional perfection as pragmatic strengthening'. *Journal of Pragmatics* **32**: 289-326.
- Householder, Fred W. (1971), *Linguistic Speculations*. Cambridge University Press. London and New York.
- Jäger, Gerhard (2000), 'Some notes on the formal properties of bidirectional optimality theory'. In Reinhard Blutner and Gerhard Jäger (eds.), *Studies in Optimality Theory*. Linguistics in Potsdam. Potsdam. 41-63.
- Kamp, H. (1975), 'Two theories about adjectives'. In Edward L. Keenan (ed.), *Formal Semantics for Natural Language*. Cambridge University Press. Cambridge. 123-155.
- Kaplan, David (1979), 'On the logic of demonstratives'. *Journal of Philosophical Logic* **8**: 81-89.
- Katz, Jerrold J. & Jerry A. Fodor (1963), 'The structure of semantic theory'. *Language* **39**: 170-210.
- Keenan, Edward L. (1974): 'The functional principle: Generalizing the notion of Subject of'. In: *Papers from the Tenth Regional Meeting of the Chicago Linguistic Society*. Chicago: Illinois, 298-310.
- Kintsch, Walter (1974), *Representation of Meaning in Memory*. Erlbaum. Hillsdale, New Jersey.
- Kiparsky, Paul. (1983), 'Word-formation and the lexicon'. In F. Ingeman (ed), *Proceedings of the 1982 Mid-America Linguistic Conference*.
- Lahav, Ran (1989): 'Against compositionality: the case of adjectives'. *Philosophical Studies* **55**: 111-129.
- Lahav, Ran (1993): 'The combinatorial-connectionist debate and the pragmatics of adjectives'. *Pragmatics and Cognition* **1**: 71-88.
- Lakoff, George (1987), *Women, fire, and dangerous things; what categories reveal about the mind*. University of Chicago Press. Chicago.
- Lascarides, Alex & Nicholas Asher, N. (1993): 'Temporal interpretation, discourse relation, and common sense entailment'. *Linguistics and Philosophy* **16**: 437-494.

- Lascarides, Alex, Ted Briscoe, Nicholas Asher & Ann Copestake (1995): 'Order independent and persistent typed default unification'. *Linguistics and Philosophy* **19**: 1-90.
- Levinson, Stephen C. (1983), *Pragmatics*. Cambridge University Press. Cambridge, England.
- Levinson, Stephen C. (1987b), 'Pragmatics and the grammar of anaphora'. *Journal of Linguistics* **23**: 379-434.
- Levinson, Stephen C. (2000a), *Presumptive Meaning*. The MIT Press. Cambridge, Mass.
- Matsumoto, Yo. (1995), 'The conversational condition on Horn scales'. *Linguistics and Philosophy* **18**: 21-60.
- McCarthy, John (1980), 'Circumscription - a form of non-monotonic reasoning', *Artificial Intelligence* **13**: 27-39.
- McCawley, James D. (1971), 'Interpretive semantics meets Frankenstein'. *Foundations of Language* **7**: 285-296.
- McCawley, James D. (1978), 'Conversational implicature and the lexicon'. In Peter Cole (ed.), *Syntax and Semantics 9: Pragmatics*. Academic Press. New York. 245-259.
- Meyer, Ralf (1993), *Compound Comprehension in Isolation and in Context*. Max Niemeyer Verlag. Tübingen.
- Merin, Arthur (1999), 'Information, relevance, and social decisionmaking: some principles and results of decision-theoretic semantics'. In Lawrence S. Moss, Jonathan Ginzburg and Maarten de Rijke (eds.), *Logic, Language, and Computation Vol II*. CSLI Publications. Stanford, Cal.
- Montague, Richard (1970), 'Universal Grammar'. *Theoria* **36**: 373-398.
- Nunberg, Geoffrey (1979), 'The non-uniqueness of semantic solutions: Polysemy'. *Linguistics and Philosophy* **3**: 143-184.
- Nunberg, Geoffrey & Annie Zaenen (1992), 'Systematic polysemy in lexicology and lexicography'. In: K. Varantola, H. Tommola, T. Salmi-Tolonen & J. Schopp (eds), *Euralex II, Tampere, Finland*.
- Partee, Barbara (1984a), 'Compositionality'. In: Fred Landman and Frank Veltman (eds.): *Varieties of Formal Semantics*. Foris. Dordrecht. 281-311.
- Prince, Alan & Paul Smolensky (1993), *Optimality Theory. Constraint Interaction in Generative Grammar*. Book manuscript Rutgers University, New Brunswick, N.J. and University of Colorado, Boulder. (To appear, MIT Press, Cambridge, Mass.).
- Pustejovsky, James (1991), 'The generative lexicon'. *Computational Linguistics* **17**: 409-441.
- Pustejovsky, James (1993), 'Type coercion and lexical selection'. In James Pustejovsky (ed.), *Semantics and the Lexicon*. Kluwer. Dordrecht. 73-96.
- Pustejovsky, James (1995): *The Generative Lexicon*. The MIT Press. Cambridge, Mass.
- Quine, Willard van Orman. (1960): *Word and Object*. MIT Press. Cambridge, Mass.
- Sag, Ivan A. (1981), 'Formal semantics and extralinguistic context'. In Peter Cole (ed), *Radical Pragmatics*. Academic Press. New York. 273-294.
- Shoham, Yoav (1988), *Reasoning about change*, MIT Press, Cambridge, Mass.
- Sperber, Dan and Deidre Wilson (1986a), *Relevance: Communication and Cognition*. Blackwell. Oxford.

- Sweetser, Eve E. (1990), *From Etymology to Pragmatics*. Cambridge University Press. Cambridge.
- Tarski, Alfred (1930), ‘Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften I’, *Monatshefte für Mathematik und Physik* **37**: 361-404.
- Tarski, Alfred (1935), ‘Grundzüge des Systemkalküls I’, *Fund. Math.* **25**: 503-526.
- Tesar, Bruce & Paul Smolensky (2000), *Learnability in Optimality Theory*. MIT Press. Cambridge, Mass.
- Van Deemter, Kees & Stanley Peters (eds.) (1996), *Semantic Ambiguity and Under-specification*. CSLI Publications. Stanford, CA.
- van Rooy, Robert (to appear a). Signalling games select Horn strategies. To appear, *Linguistics and Philosophy*.
- van Rooy, Robert (to appear b). Comparing questions and answers: A bit of logic, a bit of language, and some bits of information. In J. Groenendijk et al. (eds.), *Sources and Streams of Information*. Stanford: CSLI.
- van Rooy, Robert. (to appear c). Relevance and bidirectional OT. In R. Blutner & H. Zeevat (eds.), *Optimality Theory and Pragmatics*. New York: Palgrave Macmillan
- Williams, Edwin (1997), ‘Blocking and anaphora’. *Linguistic Inquiry* **28**: 577-628.
- Wu, Dejai (1990), ‘Probabilistic unification-based intergration of syntactic and semantic preferences for nominal compounds’. In *Proceedings of the 13th International Conference on Computational Linguistics (COLING 90)*. Helsinki. 413-418.
- Wurzel, Wolfgang U. (1998), ‘On markedness’. *Theoretical Linguistics* **24**: 53-71.
- Zeevat, Henk (1999a), ‘Explaining presupposition triggers’. Manuscript AC99. University of Amsterdam. Available from <http://www.hum.uva.nl/computerlinguistiek/henk/>
- Zeevat, Henk (1999b), ‘Semantics in optimality theory’. In Helen de Hoop and Henriëtte de Swart (eds.), *Papers on Optimality Theoretic Semantics*. Utrecht University of Linguistics. Utrecht. 76-87.
- Zimmermann, T. Ede (1991), ‘Kontextabhängigkeit’. In Dieter Wunderlich and Arnim von Stechow (eds.), *Semantik. Ein internationales Handbuch der zeitgenössischen Forschung*. De Gruyter. Berlin. 156-229.
- Zipf, George K. (1949), *Human Behavior and the Principle of Least Effort*. Addison-Wesley. Cambridge.

Endnotes

1. This is arguable, depending on the nature and depth of the syntax involved; note that they are discriminated by passive: *The {lion#bus} was taken back to the zoo*. In Relational Grammar, *the lion* is an initial 2 but *the bus* is not.
2. According to Tarski (1930, 1935), a logical consequence relation \models has to satisfy the following principles (here Γ and Γ' range over sets of formulas and ϕ over isolated formulas of a formal language L):
 - a. REFLEXIVITY: $\Gamma \models \Gamma$
 - b. CUT: if $\Gamma \models \Gamma'$ and $\Gamma \cup \Gamma' \models \phi$, then $\Gamma \models \phi$
 - c. MONOTONICITY: if $\Gamma \models \phi$, then $\Gamma \cup \Gamma' \models \phi$

-
3. A perfect number is a natural number that is identical to the sum of its true divisors; e.g. $6 = 1+2+3$ or $28 = 1+2+4+7+14$.
 4. The non-monotonic system I have in mind corresponds to the so called VORONOI TESSELATION defining a partitioning of some (abstract) space in terms of a given set of prototypes. The construction stipulates that the element x belongs to the same category as the closed prototype of the given set of prototypes. It is evident that previously defined categories may change when we add new prototypes. (For more details and for the cognitive significance of this construction, see Gärdenfors 2000.) The example may also be used for demonstrating that it is not the notion of decomposition per se that leads to the non-monotonicity of the system. This results from the fact that we define prototypes in terms of certain (binary or continuous) features.
 5. Again, it is the classical, deductive character of the entailment relation that leads to this conclusion.
 6. Unfortunately, the term “systematic polysemy” covers a whole family of empirically different subphenomena for which no unified terminology is available. Expressions such as OPEN AND CLOSED POLYSEMY (Deane 1988), CONCEPTUAL SPECIFICATION and CONCEPTUAL SHIFT (Bierwisch 1983), SENSE MODULATION and SENSE CHANGE (Cruse 1986), CONSTRUCTIONAL POLYSEMY and SENSE EXTENSION (Copstake & Briscoe 1995) may be convenient for indicating a rough outline of the classification.
 7. This contrasts with the view of Geis & Zwicky (1971) who insist that their “invited inferences” are notably different from “conversational implicatures” in the sense of Grice.
 8. For excellent discussions of the phenomenon of negative strengthening see Horn (1989: Chap. 5) and Levinson (2000a).
 9. In this place, it is not even possible to give a rough outline of this theory. Therefore, I refer to the original literature, e.g. Kaplan (1979), and to a review article by Zimmermann (1991).
 10. Bierwisch’s account can be seen as the representational counterpart of the classical theory of context-dependency.
 11. Roughly, this cost is correlated with the surprise that the particular enrichment m has for an agent confronted with the underspecified representation f .
 12. Obviously, the cost function can be used to define the ordering, but not vice versa. The connection, of course, is as follows: $\langle f, m \rangle \succ \langle f, n \rangle$ iff $\underline{c}(f, m) < \underline{c}(f, n)$.
 13. Being more pedantic, we should write \succ_{\circ} in order to indicate the dependence on the actual context \circ . We can drop the index because here and in the following we assume the actual context to be fixed.
 14. In fact the ideas of Horn, Atlas, and Levinson were the origin inspiration for developing my version of bidirectional OT.
 15. It should be noted that Dekker & van Rooy (1999) give bidirectional OT a game theoretic interpretation where the optimal pairs can be characterized as so-called NASH EQUILIBRIA.
 16. In the diagram, optimal *peps* are marked \mathfrak{S} ; the remaining super-optimal pairs are marked by \mathfrak{S}^* .
 17. The recursive notation of bidirection accounts for the interaction between Q and I/R that is informally expressed already in Horn (1984a). The advantage of the present formalization is that it allows us to PROVE the general pattern of iconicity (subsuming Horn’s division of pragmatic labour; Wurzel’s (1998) constructional iconicity and Levinson’s (2000a) M-principle – the latter expressing the second half of the pattern only.)
 18. In my opinion, Levinson (2000a) tries to turn a plausible heuristic classification scheme based on the three principles Q, I, and M into a general theory by stipulating a ranking $Q > M > I$. Accepting the heuristic classification schema, I see problems for this theory, which is burdened with too many stipulations. Not unlike Horn’s conception, I prefer to see the M-principle as an epiphenomenon that results from the interaction of Zipf’s two “economy principles” (Q and R in Horn’s terminology).