POLARITY SENSITIVITY AS LEXICAL SEMANTICS

0. PRELIMINARY

Over the last thirty years, the phenomenon of polarity sensitivity has proven both a touchstone and a stumbling block for theories of grammatical representation. Unfortunately, an abundance of scrutiny does not always guarantee an increase in insight. Two major pitfalls are worth mentioning. On the one hand, as the theorist strives for intimations of universality, the complexity and subtle variability of the data are easily underestimated or ignored. On the other hand, when one considers the phenomenon in all its glorious messiness, one may quickly despair of ever finding any general explanation.

This paper seeks to negotiate these dangers by considering polarity sensitivity as a problem in lexical semantics. The basic strategy builds on recent analyses by Krifka (1990, 1994), Kadmon and Landman (1993), and Lee and Horn (1994), all of which offer lexical semantic explanations for the distribution of polarity sensitive items (PSIs). The goal is to discover what sorts of general properties, beyond their common distributional sensitivities, might unite the large and apparently heterogeneous class of polarity sensitive items. I will argue that such properties are readily available for most, perhaps even all PSIs, and further that the distribution of these forms does indeed arise as a natural consequence of these properties. In particular I suggest that polarity items are conventionally specified for two scalar semantic features, quantitative value and informative value, and that the interaction of these two features in a single lexical form is what creates the effect of polarity sensitivity.

The account developed here draws from a large literature on the roles of semantics and pragmatics in negative polarity licensing (see in particular Ladusaw, 1980, 1983 and Linebarger, 1980, 1987, 1991), ultimately going back to the work of Horn (1972) and Fauconnier (1975a, 1975b, 1980) on polarity and pragmatic scales. Fauconnier held that polarity phenomena are not simply a matter of linguistic representations, but reflect the importance of scalar reasoning as an element of conceptual structure. I concur. I hold that the acceptability of a PSI in a given sentence is determined by the informational value, in context, of the proposition to which the PSI contributes its meaning. PSIs are understood as scalar

operators which must be interpreted with respect to an appropriately structured scalar model: they are forms whose lexical semantic-pragmatic content makes them sensitive to scalar inferences. The proposal thus departs from accounts of licensing based on syntactic configurations (Klima, 1964; Progovac, 1992, 1994; Laka, 1990; Uribe-Etxebarria, 1994) and also, though more subtly, from those based on semantic entailment (Ladusaw, 1980; van der Wouden, 1994). Ultimately, I suggest that the grammar of polarity sensitivity is based not just on syntax or semantics, but crucially on pragmatic factors which determine what one may reasonably infer from the use in context of a given proposition.

Let me be clear at the outset about the limitations of my proposal. PSIs vary widely both within and across languages and yet an adequate analysis for even a single one of these forms can be surprisingly elusive. In this paper I can offer only a broad account of what all these forms might plausibly share and suggest some major lines along which they might vary. My goal is not to solve all the puzzles of polarity sensitivity, just to unite them as facets of one general problem. Moreover, while I seek to explain polarity sensitivity in terms of lexical semantic properties which PSIs encode, I can offer no way of predicting what forms will have these properties and so count as polarity sensitive. The properties I suggest are independently motivated and commonplace semantic constructs, but as with all semantic properties, their association with a given form is arbitrary and may not be detectable independently of the distributional behavior they trigger. The point is that these peculiar behaviors themselves are not arbitrary and need not be independently stipulated of the forms which exhibit them: polarity items can be listed directly in terms of their semantic content and without any formal stipulations. Of course, this is not a gain in economy: a distributional stipulation is simply replaced by a lexical semantic one; however, I hope that the account here can offer at least some insight into the basic mystery of polarity items, namely why they should exist in the first place. Polarity items exist because they are useful, because the distinctions they encode and which make them polarity sensitive serve basic semantic and pragmatic functions.

1. THREE PROBLEMS OF POLARITY SENSITIVITY

Polarity sensitivity is essentially a distributional phenomenon: in many languages, certain lexical items are sensitive to the polarity (positive or negative) of the sentences in which they appear. In English, for example, the negative polarity items (NPIs) *at all* and *much* are fine with sentential

negation as in (1a) and (2a), but unacceptable in simple affirmative sentences like (1b) and (2b).

(1) a. Sally didn't like the marzipan at all.
   b. *Sally liked the marzipan at all.
(2) a. Albert didn't get much sleep.
   b. *Albert got much sleep.

Conversely, the positive polarity items (PPIs) *sorta* and postmodifying *as hell* are fine in simple affirmative sentences like (3b) and (4b) but unacceptable with sentential negation as in (3a) and (4a).

(3) a. *Maggie wasn't sorta rude to her secretary.
   b. Maggie was sorta rude to her secretary.
(4) a. *Bert wasn't rude as hell to Ernie.
   b. Bert was rude as hell to Ernie.[1]

These facts would seem to suggest a simple syntactic explanation in which the acceptability of polarity items is conditioned by the presence or absence of an overt negation in the sentence.

But the problem is considerably more interesting than that. PSIs turn out to be sensitive to a wide range of contexts beyond simple sentential negation. These contexts include, but are not limited to, questions, comparatives, conditionals, the complements of factive adversatives, relative clauses headed by universal quantifiers, the subordinators *before* and *long after*, certain VP-adverbs such as *seldom*, *rarely* and *hardly*, the determiner *few*, and the scopal adverb *only*. A small but representative sample of these contexts is given in (5)–(9): in the a-sentences a negative polarity trigger, represented in uppercase letters, licenses the NPI *at all* and blocks the PPI *sorta*; in the b-sentences the trigger is absent and so the PPI is licensed while the NPI is unacceptable.

(5) a. FEW of the guests were (at all/*sorta) rude.
   b. Some of the guests were (*at all/sorta) rude.

---

[1] As Baker points out, these facts hold "with normal intonation and no special context" (1970: 169). With metalinguistic negation, for example, PPIs may be acceptable and NPIs will not be (cf. Linebarger, 1980; Horn, 1985). A positive sentence used to contradict a previous negative assertion will exclude PPIs and will sometimes allow some NPIs, but generally only those which have some chance of being used jocularly in simple affirmatives (i.e. *a shred of evidence* but not *any* or *ever*). In general, PPIs often seem less sensitive than NPIs: their behavior may be less constrained and judgements about them are usually less robust. Elaborating on a suggestion of Horn's (p.c.), this asymmetry may be due to the fact that while the conditioning factors for NPIs are overt, those for PPIs are not. Put simply, it may be easier to notice that something is present than to notice that something is absent, and so positive constraints may in general be more robust than negative ones.

(6) a.   ONLY Herbert was (at all/*?sorta) impressed by the ice-danc-
         ing.
     b.   Even Herbert was (*at all/sorta) impressed by the ice-dancing.
(7) a.   EVERYONE WHO was (at all/*?sorta) patriotic was wildly
         waving their flag.
     b.   Many of the onlookers who were (*at all/sorta) patriotic were
         wildly waving their flags.
(8) a.   IF Gwen is (at all/?sorta) late, she is going to be grounded.
     b.   Because Gwen was (*at all/sorta) late, she was grounded.
(9) a.   I'm AMAZED that Herbert was (at all/*sorta) interested in
         birdwatching.
     b.   I knew that Herbert was (*at all/sorta) interested in birdwatch-
         ing.

As these examples clearly show, whatever it is that PSIs are sensitive to,
it's not just negation. The first problem of polarity sensitivity then is to
find some way of characterizing the diverse array of licensing contexts as
a natural class. Klima (1964) first addressed this problem by stipulating
that environments which license NPIs share a feature, [+Affective], which
non-licensing environments lack. Since then, the goal has been to give
some substance to this notion of affectivity that would explain why it
licenses NPIs. This is the licensing problem.

Solutions to the licensing problem are usefully divided into those which
are basically syntactic and those which are basically semantic. Syntactic
approaches tend to assume an overt negative form in a specific structural
position as a primary licensing mechanism (Jackendoff, 1969; Baker, 1972;
Linebarger, 1980, 1987, 1991; Progovac, 1988, 1992; Laka, 1990; Uribe-
Etxebarria, 1994). Any residue of "non-negative polarity licensing" is
then handled by secondary semantic or pragmatic principles. Semantic
approaches, on the other hand, view negation as just one licenser among
many, and so seek general logical or pragmatic principles that can unite
them all (Fauconnier, 1975a, b, 1979, 1980; Ladusaw, 1980, 1983; Hoek-
sema, 1983; Krifka, 1994; Zwarts, 1990; Kadmon and Landman, 1993;
Lee and Horn, 1994). Roughly, these approaches hold that licensing is
based on what sorts of inferences the licensing environment supports.

Licensing, however, is only one of many puzzles. Assuming there is
some general feature that unites the diverse licensing contexts, we will
still want to know why it is that PSIs are sensitive to just this particular
feature. This problem, the sensitivity problem, is really just the lexical
semantic mirror of the licensing problem: while the licensing problem asks

why certain contexts trigger polarity sensitivity, the sensitivity problem asks what makes certain forms so sensitive to these contexts.

Logically, the two problems go together and an adequate solution to the one will hopefully provide a basis for solving the other. Granted, it makes good methodological sense to treat the licensing problem as primary. PSIs are defined as a class on the basis of syntactic distributions, and so it is natural to start the search for whatever it is that makes PSIs special by examining those syntactic distributions. Unfortunately, the assumption has often been that the sensitivity problem is not only methodologically secondary, but theoretically insignificant as well.

This general insensitivity to the sensitivity problem may be rooted in a common theoretical prejudice holding that grammatical phenomena are arbitrary and unaffected by considerations of meaning – a prejudice which makes it reasonable to think that the sorts of distributional generalizations that explain the licensing problem will in principle be independent from any lexical semantic considerations that could explain sensitivity. Of course, one only finds the generalizations one looks for, and if there are lexical semantic generalizations to be found, they may well have important implications for a theory of polarity sensitivity. Recent work by Krifka (1990, 1994), Kadmon and Landman (1993), and Lee and Horn (1994) has laid the basis for a lexical semantic approach to polarity sensitivity. These works have sought plausible lexical semantic features which might help explain the distributional behavior of certain classes of PSIs. The present paper builds on the insights of these earlier works by providing a more comprehensive view of polarity sensitive phenomena.

Although the licensing and sensitivity problems are the crucial explicanda for a theory of polarity sensitivity (and the focus of this paper), a full account will also have to deal with the diversity problem. The range of items which count as PSIs is at least as broad as the range of contexts which license them, and their variation, both cross- and intra-linguistically, is breathtaking. Within a given language PSIs may serve a variety of semantic, pragmatic and grammatical functions. In English alone the set of PSIs includes indefinite determiners, aspectual adverbs, auxiliary verbs, conjunctions, VP idioms and a variety of adverbial intensifiers (for an extensive overview, cf. von Bergen and von Bergen, 1993). Moreover, PSIs fulfilling equivalent roles in different languages may vary widely both in their morphology and their precise distribution. As Haspelmath (1993) has shown for the indefinites, such variation is both more complex and more systematic than one might expect.

Not surprisingly, different PSIs, both within and across languages, often show distinct patterns of sensitivity. The problem is particularly well docu-

mented with the NPIs. Some, like the indefinites *any* and *ever*, occur in basically all licensing environments; others, like punctual *until*, are more particular about which licensers they allow; and some, like certain romance N-words and Serbo-Croatian NI-NPIs, require an overt negation to be licensed (cf. Progovac, 1994). It is usually assumed that such differences in sensitivity reflect the relative strength of different NPIs, with some NPIs requiring the strong licenser of an overt negation and others accepting various weaker licensers (Horn, 1970; Linebarger, 1980; Edmondson, 1981, 1983; Hoeksema, 1983; van der Wouden, 1994); however, as noted in Israel (1995a, b), it is not clear that NPIs can be neatly ordered from weak to strong, nor that the diverse range of triggers can be reduced to a one dimensional gradient of licensing power. Ultimately, any adequate account of the diversity problem will have to face the fact that different classes of PSIs may require rather distinct sorts of explanation. Indeed, given the importance of lexical factors such as fossilization and collocationality in creating PSIs (van der Wouden, 1994; Hoeksema, 1994), it seems that in fact every PSI may have its own story.

Much work on polarity has neglected this diversity, blinded, as it were, by the desire for universal principles of grammar. Still, a healthy respect for diversity need not preclude a search for grand overarching patterns. In what follows I develop a proposal which, while focused on English, is intended to extend naturally to other languages and to accommodate a comprehensive range of PSIs. Starting with the sensitivity problem, I argue in Section 2 that polarity sensitivity in general arises from the interaction of two sorts of lexical semantic properties: PSIs are lexical expressions combining a high or a low quantitative value with a conventionally emphatic or understating informative value. In Section 3, I sharpen these observations by arguing that PSIs are scalar operators whose interpretation is linked to the availability of an appropriately structured scalar model (Fillmore, Kay and O'Connor, 1988; Kay, 1990). In Sections 4 and 5, I compare my analysis to some alternative approaches, arguing that with no sacrifice in explanatory elegance the scalar model account achieves greater empirical coverage than has heretofore been possible. In Section 6, I conclude by considering some of what remains to be done for a complete theory of polarity sensitivity, along with some speculations on how it might get accomplished.

2. The Lexical Semantics of Polarity Sensitivity

In this section I argue that polarity sensitivity arises from the interaction of two binary lexical semantic features: (quantitative) q-value, which can

be either high or low, and (informative) i-value, which can be either emphatic or understating. Quantitative value simply refers to an element's position within a scalar ordering and reflects the well-known fact that a sizable portion of PSIs encode some notion of amount or degree. The notion of informative value (cf. Kay, 1990) reflects the fact that in context and with respect to background expectations some propositions are more informative than others: moreover, in characterizing any given situation, a speaker may exploit this fact to present her contribution either as strongly informative and emphatic, or as weakly informative and understating. As I will argue, both these features are independently motivated and play an important role in the lexical semantics of non-PSIs. With respect to PSIs, the two features define a taxonomy of four classes distinguished on the basis of lexical semantics, each of which is amply represented in English and other languages, and each of which is characterized by distinct semantic and distributional properties.

2.1. *The Four Sorts of Polarity Sensitive Items*

The basic descriptive observation elaborated below is that PSIs consistently both designate a high or a low q-value, and are conventionally associated with an emphatic or an understating i-value. I argue that polarity sensitivity is causally linked to these two features in such a way that a form's being specified for both features is a sufficient and perhaps a necessary condition for its being polarity sensitive. In Section 3 I return to the basic notions of q-value and i-value, providing more precise definitions for both in terms of the structure of a scalar model.

The contrast in 10 between the NPIs *much* and *a wink* illustrates how the features work.

(10) a. Margo didn't sleep a wink before her big test.
     b. Margo didn't sleep much before her big test.

Intuitively, the difference between these sentences is obvious: (10a) makes a strong claim by denying that Margo slept even the smallest amount imaginable; (10b) makes a weak claim by denying only that Margo slept for a long time. In (10a), *a wink* marks a low, in fact a minimal, quantita-

tive value and produces an emphatic sentence; in (10b), *much* marks a relatively high quantitative value and produces an understatement.[2]

Similar examples abound. As many linguists have noted, expressions denoting minimal quantities or scalar endpoints often become stereotyped as emphatic NPIs (Borkin, 1971; Schmerling, 1971; Fauconnier, 1975a). Examples in English include *drink a drop*, *(spend) a red cent*, *budge an inch*, *lift a finger* and *have a snowball's chance in hell*, to name a very few. Further examples are found in languages as diverse as Sanskrit, French, Irish, Maltese, Lezgian, Dutch, Persian, Basque and Japanese (for references and examples, see Horn, 1989, p. 452 and Haspelmath, 1993, pp. 220–222). Other emphatic NPIs include the scalar conjunctions *let alone* and *much less*, degree adverbs like *at all*, *in the slightest*, and *the least bit*, and a variety of verbs and verbal idioms such as *budge*, *can stomach*, *can fathom*, *would dream of* and *can possibly*. Also included in this class are the classic indefinite polarity items *any* and *ever*, which in most, though not all (cf. Rullmann, 1996), of their uses are clearly emphatic. Differences between indefinite and minimizer NPIs are discussed in Israel (1995a).

Understating NPIs patterning like *much* in (10b) are somewhat less common, but they do constitute a clear natural class. Other examples in English include the temporal adverbial *long*, as in "He didn't last long"; *all that* +Adj, as in "Few of them are really all that clever"; and certain uses of many, which in colloquial speech tends to be replaced by *a lot of* in positive contexts. Examples from other languages include the French *grande chose*, 'a whole lot', the Dutch NPIs *pluis*, literally, 'plush,' roughly, 'problem-free,' and *mals*, 'tender, gentle,' and the Persian NPIs *cœndan*, 'much' and *un-qœdrha*, 'that much' (for discussion of Dutch NPIs see van der Wouden (1994); for Persian, see Raghibdoust (1995)).

Appropriately enough, everything is backwards when polarity is reversed: the neat division of NPIs into low scalar emphatics and high scalar understaters is neatly mirrored by a division of PPIs into high scalar

---

[2] The distinction between a hedge and an understatement is not crucial here, but it is none the less real. More or less following Hübler (1983), we can distinguish the two as different strategies of saying less than one means. In understatements it is the content of a claim that is minimized, whereas in a hedge it is the speaker's commitment to the claim that is minimized. Thus, with respect to a proposition like "Stella is very clever," (i) would be an understatement, while (ii) would be a hedge.

(i)   Stella is fairly clever.
(ii)   I guess Stella is clever.

As Kay (1983) points out, forms like *sorta* and *kinda*, among many others, may convey either an understatement or a hedge. (cf. also Bolinger, 1972; G. Lakoff, 1972.)

emphatics and low scalar understaters. Consider the contrast between the low-scalar PPI, *a little bit* and the high-scalar, *scads*. (The status of these expressions as PPIs is demonstrated by their unacceptability with the NPI trigger *rarely*.)

(11) a. Belinda (*rarely) won scads of money at the Blackjack tables.
     b. Belinda (*rarely) won a little bit of money at the Blackjack tables.

Once again, the difference is intuitively straightforward: (11a) constitutes an emphatic assertion to the effect that Belinda won a very large quantity of money, while (11b) modestly asserts only that Belinda won (at least) a small quantity of money. Once again, there is a correlation between a polarity item's informative and quantitative values, only here the correlation is the mirror image of what we found with the NPIs in (10): *scads* designates a very high quantity and produces an emphatic sentence; *a little bit* designates a small quantity and produces an understatement.

Similar examples of both low-scalar hedging and high-scalar emphatic PPIs are readily multiplied. Low-scalar PPIs in English include weak referential indefinites like *some*, degree modifiers such as *pretty*, *rather*, *somewhat*, and *sorta*, VP-idioms like *give X a shot* and *put in a word for*, and quantificational NPs like *a mite*, *a smidgen*, *a tad* and *a handful*. Examples from other languages include the French *plutôt*, 'rather;' the Dutch *een beetje*, 'a bit' (cf. van der Wouden, 1994, p. 51); and Persian forms like *qœdri*, 'a bit', *kœm kœm*, 'little by little,' and the idiomatic VP *ye qolop xordœn*, 'to drink a gulp' (examples from Raghibdoust, 1994).

High-scalar PPIs, what Hinds (1974) called "doubleplusgood polarity items," include comparative and superlative expressions such as *far Xer*, *way Xer* and *by far the Xest*, intensifiers such as *utterly*, *awfully*, *damnably*, *entirely*, *intensely* and *as hell*; quantifying NPs such as *heaps*, *mountains* and *tons*, universalizing idioms like *all the time in the world*, *all smiles* and *the whole kit and caboodle*, and a large class of slangy and unstable evaluative adjectives such as (in some registers of my idiolect) *bitchin*, *awesome*, *radical*, *gnarly* and *way cool*. Examples can be multiplied almost endlessly from any language. Van Os (1989) suggests that in German most intensifiers are PPIs (cited by van der Wouden, 1994, p. 12), and van der Wouden himself suggests that in any language most, if not all, "inherently intensified" lexical items will be PPIs (p. 19).

The lexicalization pattern of PPIs mirrors that of NPIs. While low scalar understaters are PPIs, low scalar emphatics are NPIs, and conversely, while high scalar understaters are NPIs, high scalar emphatics are PPIs.
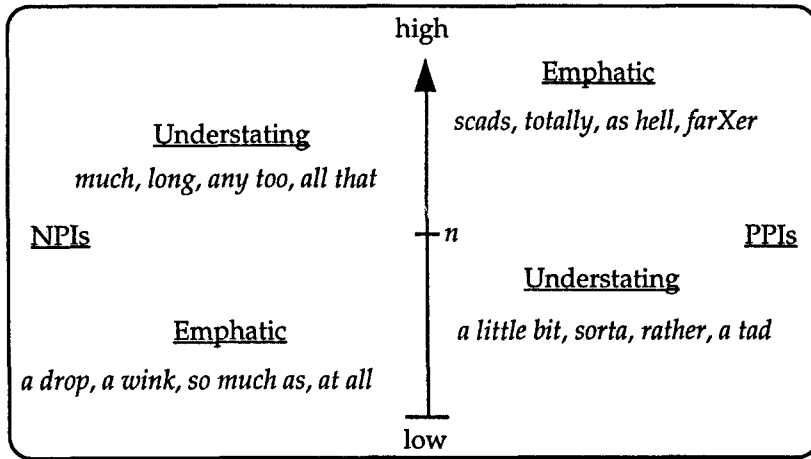
Fig. 1.

This situation is depicted schematically in Figure 1, which presents the four sorts of PSI arranged in terms of their quantitative and informative values. Note that quantitative value need not be absolute but is in fact often understood as relative to some scalar norm, represented as n in the diagram. Furthermore, while emphatic PSIs tend to mark extreme q-values, lying at or near a scalar endpoint, understaters tend to lie in the middle of the scale, clustering around the scalar norm.

Before proceeding to a more detailed defense of the taxonomy, I might just point out how thoroughly banal it is. All four of these lexical classes have, in one way or another, been identified and discussed in the literature. One of these classes, the low-scalar emphatics, has served as a stereotypical source for examples of NPIs (cf. Borkin, 1971; Schmerling, 1971; Fauconnier, 1975a, 1980; Heim, 1984). At the same time, the formation of understatements via the denial of high scalar expressions has received its fair share of attention in studies of emphasis, understatement and intensification (Spitzbardt, 1963; Bolinger, 1972; Hübler, 1983; Horn, 1989); but these studies tend not to focus on the polarity sensitive nature of the emphatic and understating forms they investigate. In work on PSIs, while Linebarger, for example, does explicitly recognize both "scalar endpoint" NPIs and "understater" NPIs as distinct classes (1980, pp. 236–7), she implicitly denies any connection between the two, claiming that each has its own distinct pragmatic motivation (1980, p. 248). On the other hand, Krifka (1990, 1994) does note a systematic correlation between high scalar PPIs and low scalar NPIs, but, he ignores the understating PSIs

altogether, effectively predicting that high scalar NPIs and low scalar PPIs shouldn't exist (see section 4.2, below).

The proposed taxonomy is thus neither daring nor original, but it does bring together a set of facts which clearly do belong together. Each of the four pieces has already, in one way or another, been independently identified and discussed in the literature. It is nonetheless (indeed, all the more) remarkable that these pieces have not been previously put together, for together they provide new insight into the mystery of polarity sensitivity.

2.2. Evidence for the Taxonomy

We have divided PSIs into four distinct groups on intuitive and distributional grounds. My goal in this section is to show that this division is not just an organizational convenience, but reflects the essential characteristics of polarity items. Figure 1 divides polarity items along three parameters according to whether they are PPIs or NPIs, high-scalar or low-scalar, and emphatic or understating. The motivation for the first of these parameters is just that different PSIs are acceptable in different environments: NPIs require a [+ Affective] context and PPIs require a neutral or [−Affective] context. But this is precisely what we want to explain. The basic claim of this paper is that the status of any given form as a PPI or an NPI will be predictable given its status along the other two parameters.

2.2.1. Quantitative Value

The second parameter, that of q-value, reflects the fact that most PSIs clearly encode a scalar semantics. Roughly, I understand a scale as an ordering of elements along some gradable dimension of semantic space. For a form to encode a specific q-value, then, it simply has to designate some relative or absolute position within such an ordering. In principle, of course, this allows for an infinite number of distinct q-values, but languages seem to be quite stingy about lexicalizing such distinctions. For the purposes of polarity items, we only need to recognize two: high q-value and low q-value, both of which are understood relative to contextual norms associated with a given dimension.

For many expressions, and for most PSIs, q-value is a transparent element of meaning. Quantifiers and degree modifiers, for example, typically just designate an abstract scalar extent or degree, often without reference to any particular dimension. Thus a PPI like *helluv* (<hell of), as in "He's helluv tall," simply signals that the predicate holds to a very

high degree, while the NPI *at all*, as in "He's not tall at all," signals that the predicate holds to a minimal degree.

For many forms q-value is narrowly tied to a specific dimension. In some cases this is straightforward: to *sleep a wink* is clearly to sleep a minimal amount. But sometimes an expression's richer lexical content can obscure the role of q-value. Words like *love* and *beautiful* involve elaborate cultural models; but they also contrast with words like *like* and *pretty* as encoding relatively high q-values on scales of positive affect and allure, respectively. Similarly verbal NPIs like *care for* and *mind* do not just denote particular mental attitudes, but are also understaters encoding relatively high q-values on scales of positive and negative affect. Thus to say one doesn't care for something – a stereotypically polite way of expressing displeasure – amounts to denying any significant positive feeling for it. Similarly, to say one wouldn't mind something – a conventionally indirect way of expressing willingness or even desire – is to deny having a particular aversion to it.

The basic idea is that forms like these, while not being simple degree words, present their designata as contrasting with an implicit, ordered set of alternative values: in other words, expressions like *love* and *care for* stand in paradigmatic opposition to similar terms ranged along a semantic scale. Note that not all words are like this: *dance* may contrast with things like *walk*, *jump* and *slither*, and *silk* may contrast with *cotton*, *wool* and *satin*, but these are not scalar oppositions. Similarly, as suggested by an anonymous reviewer, *to buy a bike* may be scalar in the sense that one could not buy less than a bike, but this predicate is not obligatorily construed in terms of greater numbers of bikes one could buy. Not so with *love* and *care for*, which must be construed in terms of alternative degrees of a type of affect.

There are certain common types of PSI which might seem to resist a scalar analysis, but in all cases I'm aware of the resistance is more apparent than real. Consider for example auxiliary NPIs meaning 'need' such as English *need*, Dutch *hoeven*, German *brauchen* and Mandarin *yòng* (Edmondson, 1983; Hoeksema, 1994). Clearly, the parallelism here cries out for a semantic explanation. Given the traditional, Aristotelian correlation between modality and quantification, such an explanation is readily available. Necessity, the modal equivalent of universality (truth in all possible worlds), involves a high (in fact maximal) q-value on a probability scale, and so an NPI like *need* appears to be a straightforward example of a high-scalar understating NPI.[3] Incidental support for this analysis comes

---

[3] The understating nature is evident in a sentence like *You needn't concern yourself about it*. Here the strict reading is that concern is not necessary, though it may still be allowed; in

from the fact that English *must* is itself a PPI, at least in the sense that it obligatorily takes wide scope over negation and other NPI triggers. As such *must* takes up where modal *need* leaves off: *mustn't* can only mean 'necessary not'; *needn't* can only mean 'not necessary'.

Further support comes from PSIs at the opposite end of the probability scale, where possibility is the modal equivalent of particularity (truth in some possible world). Here we find epistemic *can* (as in *you can*(n't) be serious*) as an emphatic, low-scalar counterpart to the high-scalar *need*, and a host of verbal NPIs – e.g. *fathom, bear, stomach* and *stand* – that require some expression of ability or possibility to be licensed (see Horn, 1972, pp. 187ff, on these "*impossible* polarity items"). The ability operator, however expressed, functions analogously to the indefinite article in a minimizer like *sleep a wink*: in both cases the effect is to preclude specific reference and so to reinforce the irrealis nature of negation. The intricacies of these forms go well beyond the scope of this paper, but the crucial point is just that there is ample evidence for viewing modal PSIs as operators on modal scales.

Another significant class of not obviously scalar PSIs are the aspectual operators *still, yet, already, anymore* and their cross-linguistic counterparts. However, in Israel (1995c) I build on the work of Löbner (1987, 1989), Michaelis (1992, 1993) and van der Auwera (1993) to suggest that these forms make crucial reference to scales of earliness and lateness. Thus, for example, I analyze a form like *still* as encoding a high q-value on a scale of lateness and thus indicating that a proposition within its scope must be understood as holding relatively late with respect to some default expectation. Similarly, *already* is held to encode a high q-value on a scale of earliness and so to designate a proposition as holding relatively early with respect to some expectation. In current work, I am extending this analysis to forms like punctual *until*, which I suggest is a low scalar emphatic NPI: it forms a maximally informative proposition by designating the lowest point on a scale of earliness at which the proposition holds. Obviously, there is much more to be said about all of these forms, and I have tried to say at least some of it elsewhere. Unfortunately, a more detailed analysis goes beyond the scope of this paper.

The important conclusion for this section is that while different PSIs may encode quantitative values in quite different ways, the generalization that PSIs consistently express some notion of quantity appears to be quite robust.

---

practice, however, such a sentence may simply be a polite way of conveying that any concern on the part of the addressee would be unwelcome.

2.2.2. *Informative Value*

I-value is perhaps the least self-evident of the three parameters, but a variety of tests suggest that emphatic PSIs do constitute a lexical class distinct from the understating PSIs. Thus certain intensifying devices allow other intensifiers but exclude hedged constructions within their scope. In (12)–(13) the emphatic a-sentences allow modification by an intensifying *literally* while the understating b-sentences resist such modification:[4]

> (12) a. Margo literally didn't sleep a wink before her big test.
>   b. *Margo literally didn't sleep much before her big test.
> (13) a. Belinda literally won scads of money at the blackjack tables.
>   b. *Belinda literally won a little bit of money at the blackjack tables.

Similarly, in (14)–(15), the emphatics, but not the understaters can be felicitously introduced by a breathless *You'll never believe it!*

> (14)   You'll never believe it!
>   a. Margo didn't sleep a wink before her big test.
>   b. ?Margo didn't sleep much before her big test.
> (15)   You'll never believe it!
>   a. Belinda won scads of money at the blackjack tables.
>   b. ?Belinda won a little bit of money at the blackjack tables.[5]

While the contexts in (12)–(15) above favor emphatics and exclude understaters, it is difficult to find contexts which favor understaters but exclude emphatics. Thus if we substitute *sorta* for *literally* in (12)–(13), or *It kinda seems to me* for *You'll never believe it* in (14)–(15), there is little difference in acceptability between the a- and the b-sentences. It seems that while it is impossible to intensify an understatement, it is perfectly feasible to hedge an emphatic utterance. The reasons for this curious fact unfortunately defy my understanding.

Finally, the distinction between the emphatic a-sentences and the hedged b-sentences is nicely illustrated in the syntactic tests used by Horn (1972, 1989) to define quantitative scales. Roughly, these tests help establish paradigmatic relations between forms ranged on a scale: coordinating conjunctions like *or at least* require that the first conjunct represent a stronger claim than the second, while *in fact* or *and what's more* require

---

[4] Similarly, *even* and *absolutely* both allow emphatics but exclude understaters in their focus.
[5] The b-sentences may be acceptable if focus stress falls somewhere besides the PSI, but this only underscores the point that the understating PSIs are barred from contributing emphatic or controversial information to a sentence.

that the second conjunct represent a stronger claim than the first. The use of these coordinators to combine emphatic and understating PSIs bears out the intuition that emphatics make stronger assertions than understaters.

(16) a. Margo didn't sleep a wink, or at least she didn't sleep much.
    b. *Margo didn't sleep much, or at least she didn't sleep a wink.
(17) a. Margo didn't sleep much, in fact she didn't sleep a wink.
    b. *Margo didn't sleep a wink, in fact she didn't sleep much.
(18) a. Belinda won scads of money, or at least she won a little bit.
    b. *Belinda won a little bit of money, or at least she won scads.
(19) a. Belinda won a little bit of money, in fact she won scads.
    b. *Belinda won scads of money, in fact she won a little bit.

The patterns of acceptability in these sentences lend support to the claim that PSIs are divided between emphatic and understating forms.

It is worth noting that the division of PSIs into emphatic and understating forms also sheds at least a little light on the diversity problem. As the above contexts suggest, different PSIs have different licensing requirements for the simple reason that they have different lexical semantics. The contrasts in (20)–(23) illustrate licensing contexts in which understating NPIs are more awkward than their emphatic counterparts. In all these cases the contrast is sharpest when the NPIs are focused and given prosodic prominence.

(20) a. Never has he drunk a drop at any of those parties.
    b. ?Never has he drunk much at any of those parties.
(21)   I'd rather be trapped in an elevator with a lecherous Martian than spend
    a. a second with that Murray.
    b. ?much time with that Murray.
(22)   Jasmine kept pestering the coach long after
    a. she had a hope in hell of getting on the team.
    b. ?she had much hope of getting on the team.
(23) a. Everyone who likes Sally at all will be there.
    b. ?Everyone who likes Sally much will be there.

In (20), the preposed negative sets up an expectation for a truly newsworthy assertion, but the expectation is frustrated by the weakly informative *much*. In (21), the effectiveness of the comparative construction depends on the magnitude of the speaker's preference for lecherous Martians over the very distasteful Murray: the minimizer *a second* emphasizes that magnitude, while the weaker *much* diminishes it. Similarly, in (22), the construction with *long after* depends on an anticipated contrast be-

tween Jasmine's likelihood of success and the intensity of her efforts: the minimizer *a hope in hell* effectively reinforces that contrast; *much* undermines it. In (23), finally, *at all* is natural as it expands the set picked out by the universal quantifier; *much*, however, diminishes that universal force, suggesting as it does that people who like Sally only a moderate amount may well be excluded.

While the above sentences illustrate contexts that prefer emphatic over understating forms, it is also possible to find contexts in which understatement is the preferred form of expression. In (24) and (25), the weakly negative *few* and *seldom* seem to prefer the more modest force of the understaters over the emphatic minimizers.

(24) a. ?Few of them spent a red cent on their outfits.
     b. Few of them spent much on their outfits.
(25) a. ?He seldom gets a wink of sleep before a performance.
     b. He seldom gets much sleep before a performance.[6]

Finally, it is worth noting that the contrast between weakly and strongly informative PSIs also helps explain the different effects they produce in questions. Many researchers have noted that minimizer NPIs in particular force a rhetorical question (Borkin, 1971; Linebarger, 1980); Hinds (1974) makes a similar point for emphatic PSIs in negative questions. The contrasts are illustrated in (26)–(27), where the a-sentences, with emphatic PSIs can only be used rhetorically, while the b-sentences, with understating PSIs, can function more as information questions.

(26) a. Did you eat a bite of the cake? (rhetorical only)
     b. Did you eat much of the cake? (info question)
(27) a. Wasn't she awfully clever? (rhetorical only)
     b. Wasn't she sorta clever? (info question)[7]

Rhetorical questions can be understood as a species of indirect speech

---

[6] One might argue that this contrast is due to a register clash between the relatively formal *few* and *seldom*, and the relatively colloquial minimizers, *a red cent* and *a wink*. This is surely correct, but in fact it begs the question, since the difference between the two registers reflects different preferred politeness strategies which can be explicated in terms of the contrast between emphatics from understaters. Roughly, formal registers tend toward deference and so are most comfortable with the open-ended nature of understatement; more colloquial registers, however, typically emphasize camaraderie and so prefer the higher speaker involvement and unabashedness of the emphatics (cf. Lakoff, 1973; Brown and Levinson, 1979).

[7] In (27) the negative question itself signals expectation of a positive response; (27a), however, with the emphatic *awfully*, is hardly even a question so much as a request for agreement, while (27b), with the more open-ended *sorta*, at least leaves room for disagreement, as well as considerable latitude concerning the degree of cleverness.

act in which a speaker, by superficially and insincerely requesting informa-
tion, actually conveys a very definite opinion. Normally, if a question is
a sincere request for information, the speaker will not want to excessively
prejudice the set of possible responses; however, that is exactly what an
emphatic PSI will do. By posing a question with reference to an extreme
value, the speaker renders one possible response extremely informative
and the other extremely uninformative: if the answer to (26a) is "no," we
learn precisely how much cake was eaten (none); if it's "yes," we know
only that at least the smallest amount possible was eaten. Such a preju-
dicial posing generates the implicature that the speaker in fact has a very
definite idea about the answer, and so the question is rhetorical. The
understating PSIs, on the other hand, allow an interlocutor more room
for negotiation, and so can be used to form simple information questions.

Although the intricacies of the diversity problem extend well beyond
the difference between emphatics and understaters (and well beyond the
scope of this paper), the basic strategy of taking seriously the subtleties
of PSIs' lexical semantics shows promise of leading to further insight into
the differences between PSIs.

2.3. The Subtle Sensitivity of Insensitive Items

Thus far, we have established a taxonomy of PSIs based on two lexical
features: quantitative and informative value. In what follows, I argue that
it is precisely the convergence of these two features on a single lexical
item that creates polarity sensitivity. Before we consider how this works,
it is worth noting that both of these features are independently motivated,
unexceptional semantic constructs and that both play a role in the seman-
tics of other lexical items. Moreover, by distinguishing the two features
we gain a natural explanation for the otherwise idiosyncratic behavior of
a variety of apparent synonyms.

If q-value and i-value really are independent lexical features characteriz-
ing PSIs, we should expect to find forms which are conventionally specified
for one feature but not the other. And we do. The degree modifiers below
all encode low q-values, but they vary with respect to i-value: only *a bit*,
unlike its near synonyms *the least bit* (NPI) and *a tad* (PPI), can occur in
both emphatic and understating contexts.

(28) a. Harry is a bit overweight.
    b. Harry is a tad overweight.
    c. *Harry is the least bit overweight.

(29) a. Harry isn't a bit overweight.
    b. *Harry isn't a tad overweight.
    c. Harry isn't the least bit overweight.

The positive sentences in (28) all make weak claims and so can function only as understatements or hedged assertions: the emphatic NPI *the least bit* cannot be accommodated. In (29), where the same q-value yields a strong scalar claim, the sentences can only count as emphatic denials: here, the understating PPI *a tad* is ruled out. But the versatile *a bit* is fine in both situations.

A similar contrast is found in (30)–(31) between the non-PSI intensifier *very* and its sometimes near-synonyms, the PPI *awfully* and the NPI *all that*.

(30) a. Lewis is very clever.
    b. Lewis is awfully clever.
    c. *Lewis is all that clever.
(31) a. Lewis isn't very clever.
    b. *Lewis isn't awfully clever.
    c. Lewis isn't all that clever.

In (30a), *very* marks a high degree of cleverness in an emphatic assertion; in (31a), *very* marks a high degree of cleverness in a hedged denial. The b- and c-sentences show that *awfully* and *all that* are not so flexible. The notion of i-value provides a simple explanation: forms specified for a particular i-value are limited to contexts supporting that value; forms not so specified are free to occur in emphatic, understating or neutral contexts. Forms like *a bit* and *very*, while sharing a q-value with their apparent synonyms, differ in that they do not encode a conventional i-value. Their distributions are consequently less constrained.

At this point one may object that the argument has turned circular.[8] While I've claimed that polarity sensitivity is predictable on the basis of i-value and q-value, it seems that in (28)–(31) the determination of i-value itself depends on a form's polarity sensitive behavior. The objection is valid, but it may miss the point. I-value cannot be predicted from lexical semantics because i-value is itself a part of lexical semantics, and so its association with any given form is arbitrary. The question is, given i-value and q-value as lexical semantic features, is that enough to predict a form's

---

[8] I am indebted to Chris Barker, Adele Goldberg, Larry Horn, Hotze Rullmann and an anonymous reviewer for alerting me to this possibility.

polarity behavior? And if so, then just what sort of a feature is this i-value anyway?

In essence, informativity is a property of sentences used in context. Emphatic sentences convey more or somehow make a stronger claim than might have been expected; understating sentences say less or make a weaker claim than might have been expected. I-value, the sentential property, becomes a feature of lexical semantics when particular words are conventionally associated with emphatic or understating contexts. In other words, if a given form occurs frequently and systematically in emphatic contexts, the form may itself be stereotyped as conveying an emphatic pragmatic force. This sort of metonymy, which Stern (1931) calls "permutation," is in fact a common source of semantic change.[9]

Moreover, the conventionalization of i-value as an aspect of lexical meaning is consistent with, and indeed exemplary of the general tendency noted by Traugott for "meanings to become increasingly situated in a speaker's subjective . . . attitude" toward what is said (1988, p. 411). This process of pragmatic strengthening is typical of early stages of grammaticalization, and i-value seems to provide a typical example. I-value is a pragmatic feature encoding a speaker's attitude toward the content she conveys: emphatic utterances express high involvement and commitment to what is said; understatements signal deference and a desire to mitigate face threatening acts. As such, i-value is an unremarkable sort of lexical-semantic feature, and though we might not be able to predict where it will show up, we should not be surprised to find evidence of it at work.[10]

But if i-value really is an independent lexical-semantic feature, we should find forms which encode a particular i-value, but which are nonetheless not polarity sensitive because they are not conventionally associated with any particular q-value. An obvious example is *even*. A variety of proposals have been made for dealing with the peculiar contribution *even* makes to a sentence (Horn, 1969; Fauconnier, 1980; Kay, 1990; Francescotti, 1995, among others), but all agree in essence that a sentence containing *even* will express a proposition which is somehow less expected or more informative than some other contextually supplied proposition.

---

[9] Compare, for instance, the tendency of connectives expressing temporal overlap to develop concessive meanings, as with English *while, still* and *yet* (Traugott and Hopper, 1991, p. 199): often the point of saying that two things are occurring together is to draw attention to their normal incompatibility (cf. *She's seven and she's studying modal logic*), and so this notion of contraverted expectation may become associated with a marker of simultaneity.

[10] The notion of informativity as a conventional element of lexical pragmatic meaning has been explored extensively in the work of Anscombre and Ducrot (1983, and elsewhere). See also Verhagen (1995) for an account of *let alone* in terms of argumentative goals rather than simple entailment relations.

Even is not polarity sensitive, occurring freely in both negative and affirmative sentences, and *even* is not linked to any fixed q-value, since both low and high-scalar expressions can occur in its focus. But *even* is sensitive to the interaction of polarity with the scalar semantics of its focus. While both *even the lowest* and *even the highest* are perfectly well-formed phrases, generally only one of the two can occur in any given context. Moreover, as (32)–(33) illustrate, their acceptability in a given context is sensitive to the context's polarity.

(32)a. Dolly can jump over even the tallest obstacle.
   b. #Dolly can jump over even the lowest obstacle.
(33)a. Dolly can't jump over even the lowest obstacle.
   b. #Dolly can't jump over even the tallest obstacle.[11]

As Fauconnier (1975a, 1975b) points out, superlatives like those in the a-sentences function like universal quantifiers. As such, these sentences represent remarkable claims and so welcome *even* as a marker of their unusual informativeness; however, when the polarity is reversed in the b-sentences, the claims become trivial and *even* sounds bizarre. Because *even* does not itself refer to any particular point on a scale, it can freely occur in both positive and negative sentences and still retain its emphatic force; however, in order to do so, the scalar semantics of its focus must be appropriate for the polarity of the sentence.[12]

Two points emerge from these facts. First, the behavior of *even*, *very* and *a bit* clearly demonstrates the independence of quantitative and informative value: *even* encodes an emphatic i-value but is neutral as to q-value; *very* and *a bit* encode high and low q-values respectively but are neutral as to i-value. On the other hand, the behavior of these three forms

---

[11] Note that these expressions are not, strictly speaking, PSIs; rather they are polarity sensitive only with respect to a given propositional context. If we change the predicate from *can jump* to *has trouble with* the pattern of acceptability reverses:

(i)   Dolly has trouble with even the lowest obstacle.
(ii)   #Dolly has trouble with even the highest obstacle.

[12] A similar phenomenon is noted by Sweetser with respect to concessive conditionals (1990, p. 134). Either a positive or a negative apodosis may allow a concessive, *even if*, reading in (i)–(ii).

(i)   Linda wouldn't marry you (even) if you were the last man on earth.
(ii)   Linda would marry you (even) if you were a monster from Mars.

However, barring any major reorganization of background assumptions, reversing the polarity in these examples effectively blocks any concessive reading in (iii)–(iv).

(iii)   Linda would marry you (*even) if you were the last man on earth.
(iv)   Linda wouldn't marry you (*even) if you were a monster from Mars.

also demonstrates that the three parameters relevant to polarity sensitivity – polarity, q-value and i-value – all interact independently of PSIs themselves. In (32)–(33) while *even* holds i-value constant, a change in sentence polarity necessitates a change in the q-value of the focus. Similarly, in (28)–(31), while *a bit* and *very* mark a constant q-value, a change in polarity brings with it a change in the sentence's i-value. The implication of these facts for a theory of polarity sensitivity should be obvious. If an expression is such that it conventionally holds constant both quantitative and informative value, that expression will only be acceptable in contexts where its quantitative and informative values are compatible.

2.4. *Sensitivity Summarized*

In this section I have proposed an analysis of the lexical semantics of polarity sensitive items, arguing that these forms are distinguished by the conjunction of two sorts of lexical feature, one encoding a particular quantitative value and the other encoding a particular informative value. Both of these have been shown to be independently motivated, natural semantic features required for the characterization of other, non-polarity sensitive forms. Moreover, it has been suggested that forms marked for only one of these features systematically fail to be polarity sensitive, while forms marked for both inevitably are polarity sensitive. In the next section, I will attempt to explain why this should be.

3. THE LICENSING PROBLEM

Having established just what polarity sensitive items are, it remains to explain why they behave in the peculiar ways they do. Why should the combination of a particular quantitative value with a particular informative value inevitably lead to polarity sensitivity? Intuitively, the answer is simple. PSIs conventionally express a certain kind of pragmatics which limits their occurrence to just those contexts that are compatible with those pragmatics. Framed in these terms, the licensing problem reduces to another, intuitively more tractable problem, namely, what is it that makes a sentence appropriately emphatic or understating.

In this section I sketch out a solution to this problem based on the notion of a scalar model (Fillmore, Kay and O'Connor, henceforth *FKO*, 1988; Kay, 1990). Scalar models represent a refinement of Fauconnier's pragmatic scales (1975a, 1975b, 1980), consisting of a set of propositions structured so as to define pragmatic entailments between them, and the account here builds on the Fauconnier's insights about polarity pheno-

mena. Scalar models afford us a simple and precise way of defining the crucial notions of quantitative and informative value, and, given those definitions, a way of predicting when a proposition will count as either emphatic or understating. PSIs will then be understood as scalar operators whose acceptability within a sentence depends on the availability of an appropriately structured scalar model.

3.1. *Emphasis and Understatement in a Scalar Model*

As discussed above, notions like emphasis and understatement are normally considered pragmatic phenomena arising from the way sentences are used. As such, they are normally associated more with whole utterances than with individual lexical items. In Section 2, however, we saw that the acceptability of a PSI within a sentence is linked to that sentence's informative value: emphatic PSIs only occur in contexts where they can be construed as emphatic; understating PSIs only occur in contexts where they can be construed as understatements. In this section I would like to explore more precisely what it means for a lexical item to encode a notion like emphasis or understatement.

Informally, the notion of propositional informativeness depends on some sort of speaker expectation. Above, I suggested that emphatic sentences are informative because they somehow say more than one might expect, and understating sentences, at least on their literal interpretation, are uninformative because they say less than one might expect, often making claims so weak as to be trivial. So in order to give some teeth to our distinction between high and low informative values, we need some way of interpreting a proposition relative to some (possibly singleton) set of alternative propositions.

To that end, I suggest that PSIs form a special class of scalar operators. The notion of a scalar operator was first introduced by FKO (1988) as a way of handling the complex semantics and pragmatics of the NPI *let alone*. Scalar operators are themselves a special class of what Kay (1989) calls contextual operators, that is, expressions whose meanings involve not only what situations they can appropriately describe, but also some notion of the situations in which they can appropriately be used. More specifically, contextual operators

are lexical items or grammatical constructions whose semantic value consists, at least in part, of instructions to find in, or impute to, the context a certain kind of information structure and to locate the information presented by the sentence within that information structure in a specified way. (Kay, 1989, p. 181)

y_4 ⊤ hardest
y_3 ●
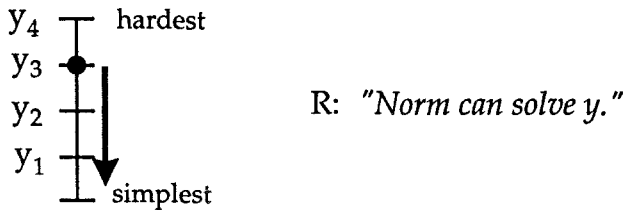y_2
y_1
⊥ simplest

R: *"Norm can solve y."*

Fig. 2.

Naturally, a contextual operator will not be acceptable if it occurs in a context where the information structures it requires can neither be found nor constructed.

The claim here is that PSIs are scalar operators: the particular sort of information structure they require is that supplied by a scalar model.[13] Informally, a scalar model consists of a structured set of propositions which a speaker and hearer either share as background knowledge or can construct in context. Propositions are understood in a standard way as functions from states of affairs to truth values, and in a scalar model they are arranged so as to support inferential relations between them. This arrangement is determined by the interaction of propositional schemata with one or more semantic dimensions. A simple, one-dimensional example might involve the propositional schema R, "Norm can solve y," and a semantic dimension D_1 consisting of the set of all puzzles ordered from the easiest to the most difficult. In general, if we know that Norm can solve a puzzle of moderate difficulty we will happily infer that he can also solve any less difficult puzzle. Within a scalar model then, whenever the propositional schema R is validly predicated of some point y_n, then R will necessarily hold for all points lower than y_n on the scale. In other words, for any two states of affairs $R(y_i)$ and $R(y_j)$, where $y_i > y_j$, $R(y_i) \geq R(y_j)$.[14] Figure 2 provides a schematic representation, where the solid arrow pointing down from y_3 represents the inferences following from the truth of $R(y_3)$.

Complicating the picture only slightly, we can apply a more open propositional schema, S, "x can solve y", to a two-dimensional scalar model pairing the set of puzzles in D_1 with a set of puzzlers, D_2, consisting of

---

[13] The notion of a scalar model goes back to the work of Horn (1972) and Fauconnier (1975a, 1975b). For detailed discussion of the formal properties of scalar models, see FKO (1988) and Kay (1990).

[14] Note that these are pragmatic entailments and are not necessarily valid outside the scalar model. Horn (1989, p. 240) notes that his quantitative scales are only "a limiting case wherein every pragmatic model or context assumes the scale in question, while other predicators are less consistent across models."
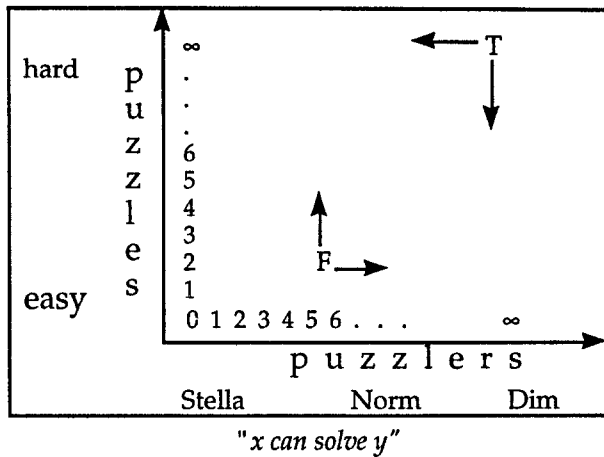
"x can solve y"

Fig. 3.

Stella, Norm and Dim. Elements along each of these two semantic dimensions will then be ordered in such a way that the validity of any higher ranked element in a given proposition will pragmatically entail the validity of any lower ranked element. Thus for a given puzzle, y, if the unusually obtuse Dim can solve y, then the more clever Norm will also be able to solve it as well. Norm will therefore be lower on the scale (alternatively, closer to the origin) than Dim. Similarly, if a given puzzler can solve a particularly difficult puzzle, say y_3, then he will also be able to solve any easier puzzle. The easiest puzzle of all will therefore be the lowest on the scale. Figure 3 (cf. Kay, 1990, p. 65) depicts this diagrammatically.

Once again, the scalar model defines a pattern of pragmatic entailments: for any proposition p within the model, if we know that p is true (i.e. has a value of T), then we can safely infer that any distinct proposition q that is lower than p on at least one dimension and no higher than p on any other dimension will also be true. Conversely, if we know that p is false (i.e. has a value of F), it follows that any proposition q that is higher than p on at least one dimension and no lower than p on any other dimension will also have a value of F. Note that certain points are inherently more informative than others: given a value of T for the point joining Stella with the easiest problem, one can infer nothing about any other proposition in the model. I will refer to this point as the **scalar origin** or the innermost proposition. A value of T for the proposition linking Dim with the hardest problem would, on the other hand, be extremely informative, entailing a value of T for every other point within the model. I will refer to this

maximally informative point as the **scalar zenith** or the outermost proposition.

Entailments in a scalar model hold relative to a given propositional schema. Figure 3 depicts entailments for the schema S, "*x can solve y*," but if we substitute the schema ~S, "*x cannot solve y*," the direction of entailments is reversed. In this case a value of T for any lower proposition p entails values of T for all propositions farther from the origin than p, while a value of F entails values of F for all propositions closer to the origin. It is thus useful to make a distinction between those schemata which, like simple affirmatives, license inferences from the zenith to the origin and those which, like negation, reverse this direction of entailments. I will call the former **entailment** (or, **scale**) **preserving** and the latter **entailment** (or, **scale**) **reversing**.

Scalar models provide a simple framework for defining the crucial notions of quantitative and informative values, allowing us to relate the two in a simple and well-motivated way. Q-value refers to a scalar operator's position in a (partially) ordered set of alternatives ranged along some dimension of a scalar model, and thus effectively determines an expressed proposition's position within the scalar model. I-value, on the other hand, refers to an expressed proposition's informativeness with respect to other propositions in the model. Kay defines informativeness by stipulating that for any two distinct propositions p and q within a scalar model p is **more informative** (or "stronger") than q iff p unilaterally entails q (1990, p. 66). This is then applied to an analysis whereby *even* is taken to mark a sentence in which it occurs as expressing "in context, a proposition which is more informative . . . than some particular distinct proposition taken to be already present in the context" (ibid.). Taking *even* as the paradigm case of an emphatic particle, we can generalize this by holding that a sentence is emphatic if and to the extent to which it expresses a proposition which is more informative than some distinct proposition available in the context. Following Kay, I will call the overtly expressed proposition the **text proposition** (tp) and the other, background proposition the **context proposition** (cp).

But how do we select a particular cp against which an expressed tp may be evaluated? I assume that in context the use of any scalar predicate evokes some scalar norm as an implicit standard of comparison, and that this scalar norm provides the cp. The particular value of the norm will depend on expectations and assumptions of the speech act participants, but in general it will simply reflect a default, real world understanding of whatever is under discussion.

An emphatic sentence, then, is one in which the expressed tp pragmati-

cally entails the proposition coded by a scalar norm. The examples in 34 are thus emphatic because, interpreted within a scalar model, they express propositions which somehow exceed normal expectations.

(34) a. Angela didn't drink a drop at that party.
     b. Huey got awfully drunk at that party.

(34a) asserts as its tp the proposition that Sally didn't drink a minimal quantity and evokes as its cp the proposition that Angela didn't drink some larger, default amount (say, two beers). Note that the cp need not reflect any particular expectation about Angela's drinking per se; rather, given that drinking is at issue, the cp simply represents what might be a normal amount for someone to drink in this context. Interpreted with respect to a scalar model pairing a dimension of drinkers with a dimension of quantities imbibed, the tp here entails the cp, and the emphatic effect reflects the disparity in strength between what is said and what might have been asserted. A similar story holds for (34b), where the asserted tp refers to an extreme state of drunkenness and contrasts with a weaker cp which would ascribe a less remarkable degree of inebriation to Huey.

Naturally, understatement turns out to be just like emphasis only backwards. An understatement is a sentence for which the overt tp is less informative (or, "weaker") than a contextually available scalar norm (the cp). With understatements then, the entailment goes from the evoked cp to the asserted tp rather than the other way around. Thus the sentences in (35) express propositions which must be understood as weaker than what one might have expected.

(35) a. Abby wasn't all that happy with her frittata.
     b. Jennifer was pretty pleased with her spinach quiche.

In (35a), the tp asserts that Abby's happiness was not particularly great, and contrasts with an evoked cp according to which her happiness was not even equal to some default norm. The evoked cp is the stronger proposition and entails the asserted tp. The effect of understatement arises from the disparity between what is asserted, the weak tp, and what might more informatively have been asserted, the stronger cp. In (35b), similarly, understatement arises from the fact that the tp picked out by

the phrase *pretty pleased* is weaker than what would have been conveyed by an unadorned *pleased*.[15]

3.2. *Toward a Solution to the Licensing Problem*

Having defined the relevant notions of emphasis and understatement in terms of scalar entailments, it remains only to determine what conditions will guarantee the right direction of entailment between a PSI's tp and its cp. The claim is that PSIs are a class of scalar operators. More precisely, a PSI is a lexical form or grammatical construction which specifies an expressed proposition p's location within a scalar model and which, by virtue of imposing a particular informative value on that proposition, further requires that p either entail or be entailed by a default context proposition q available within the scalar model.[16] The requirement that a PSI encode either an emphatic or an understating informative value thus reduces to a requirement on the direction of entailments within a scalar model. A low scalar emphatic PSI must define a proposition as occupying a low point within a scalar model and at the same time as entailing a higher cp within the model. Similarly, a high scalar understating PSI must define a proposition as occupying some high point within the model and at the same time as being entailed by a default cp closer to the origin. Thus both low scalar emphatics and high scalar understaters require a scalar model in which lower propositions entail higher ones, and both are thus negative polarity items. Conversely, high scalar emphatics and low

---

[15] It is interesting to note that the understating sentences in (35) are subject to a sort of ambiguity from which the emphatic sentences in (34) appear immune. On one reading a sentence like (35b) conveys that Jennifer's pleasure was only moderate: let us call this weak understatement. On another reading, which we can call strong understatement, the same sentence conveys that her pleasure was in fact unusually great. In the first case *pretty* functions as a detensifier, hedging the strength of *pleased*; in the second case it functions as an intensifier, reinforcing the strength of *pleased*. On the scalar model account, we can capture this distinction between strong and weak understatement by appealing to the status of the evoked cp: in weak understatement, the cp is evoked as a stronger proposition which could have been asserted, but to which the speaker is not committed; in strong understatement, the cp is evoked as a stronger proposition which does in fact hold, but which, for reasons of politeness or rhetorical effect, a speaker declines to assert (cf. Brown and Levinson on understatement as a strategy for mitigating face threatening acts, 1978, pp. 222–24). Since with emphatics the tp must by definition entail the cp, this sort of ambiguity cannot arise.

[16] This is in full agreement with Lee and Horn's (1994) claim that *any* represents an indefinite article with an incorporated *even*. Lee and Horn, harking back to Fauconnier (1975a), argue that polarity sensitive and free choice *any* are different aspects of the same scalar operator. The present paper basically generalizes this result by arguing that all (or maybe just most) PSIs are also scalar operators.

scalar understaters require scalar models in which higher propositions entail lower ones, so they are both positive polarity items.

The prediction then is that NPIs will be licensed and PPIs blocked in just those environments that reverse the direction of entailments in a scalar model; conversely, in environments that preserve the direction of entailments, NPIs will be blocked and PPIs will be licensed. In essence, this point (at least as it applies to NPIs) has already been made by Fauconnier (1975a, 1980) in terms of pragmatic scales and has since formed the basis, in one way or another, for almost every major account of NPI licensing in the literature (Ladusaw, 1980, 1983; Edmondson, 1981, 1983; Hoeksema, 1983; Heim, 1984; Progovac, 1992, 1994; Krifka, 1990, 1994; Zwarts, 1990; Kadmon and Landman, 1993; Lee and Horn, 1994; van der Wouden, 1994). Nonetheless, it won't hurt to give a few examples of how NPI-licensing environments do reverse entailments while minimally different, non-licensing environments do not.

Consider again the scalar model pairing puzzles with puzzlers in Figure 3. As pointed out above, while the propositional schema S, "*x can solve y*," is scale preserving, allowing inferences from the zenith down toward the origin, its negation ~S, "*x cannot solve y*," is scale reversing, allowing inferences from the origin to the zenith. Similar transformations of S involving the insertion of NPI licensers yield similar entailment reversals quite consistently. In (36)–(41), the uppercase forms in the a-sentences are scale reversers, as demonstrated by the pragmatic entailments from (low scalar) easy problems to (high scalar) hard problems; the minimally different forms in the b-sentences are scale preserving, as demonstrated by the pragmatic entailments from hard problems to easy problems.

(36) a.　FEW students can solve the easy problems.
　　　　　→ Few students can solve the hard problems.
　　 b.　A few students can solve the hard problems.
　　　　　→ A few students can solve the easy problems.
(37) a.　Dim can RARELY solve the easy problems.
　　　　　→ Dim can rarely solve the hard problems.
　　 b.　Dim can often solve the hard problems.
　　　　　→ Dim can often solve the easy problems.
(38) a.　EVERYONE who could solve the easy problems got some cake.
　　　　　→ Everyone who could solve the hard problems got some cake.

(38)b.   Someone who could solve the hard problems got some cake.
           → Someone who could solve the easy problems got some
             cake.
(39)a.   I'd be SURPRISED if Dim could solve the easy problems.
           → I'd be amazed if Dim could solve the hard problems.
     b.   I expected that Dim could solve the hard problems.
           → I expected that Dim could solve the easy problems.
(40)a.   ONLY Stella can solve the easy problems.
           → Only Stella can solve the hard problems.
     b.   Stella also can solve the hard problems.
           → Stella also can solve the easy problems.
(41)a.   IF Norm can solve the easy problems, he'll get some cake.
           → IF Norm can solve the hard problems, he'll get some cake.
     b.   Norm can solve the hard problems and he'll get some cake.
           → Norm can solve the easy problems and he'll get some
             cake.

As predicted, all of the NPI-licensing environments in (36)–(41) share the property of reversing entailments within a scalar model.[17]

Thus far then we have addressed the sensitivity problem by defining PSIs as a special class of scalar operators which encode both a proposition's location within a scalar model and a proposition's rhetorical informativity. Given this, and given an understanding of informative value in terms of entailment within a scalar model, the peculiar distribution of PSIs turns out to be mostly trivial. The licensing problem is solved by recognizing that PSIs require a contextually available scalar model with an appropriate direction of entailments: [+Affective] environments are just those that allow outward inferencing within a scalar model; [−Affective] environments are those that require inward inferencing within a scalar model.

Of course, much remains to be said about licensing. One important issue is the relationship between the sorts of scalar inferencing at work in (36)–(41) and the more strictly logical notion of downward entailment used in Ladusaw (1980). As discussed below, while the two notions are clearly related, they are also distinct. Other, more specific puzzles which I can only just mention include the questions of why and under what

---

[17] Note that many of these inferences depend on the relevant operators, as Larry Horn puts it (p.c.) "entailing downward even unto nothingness." Thus I assume that *only Bill can solve the hard problems* makes a true assertion even if even Bill cannot solve the hard problems (cf. Horn, 1969, 1995) and that *few students can solve the hard problems* is true even if no student can solve the hard problems. For psycholinguistic evidence that people do make these sorts of inferences in processing such operators, see Moxey and Sanford (1993, 1994).

conditions questions can license NPIs; why certain liberal NPIs are acceptable in the restriction of *only* (cf. Horn, 1996, for discussion); and why certain operators intervening between an NPI and its trigger can block acceptability. In general, to really solve the licensing problem we will require detailed analyses of the inferential properties of a wide range of subtly different polarity triggers. Moreover, as noted above with reference to the diversity problem, many PSIs show subtly idiosyncratic behavior with respect to their potential licensers. Sometimes such differences may reflect important new generalizations. For instance, Giannakidou (1994) offers evidence that certain Greek NPIs are sensitive not so much to monotonicity but rather to the veridicality of a licensing context. In other cases, however, it may well be that, as van der Wouden suggests (1994), PSIs sometimes simply have their own idiomatic collocational preferences.

Scalar models do not solve all the problems of polarity sensitivity. An abundance of mysteries remain. Still, a view of PSIs as basically constrained by the scalar inferential properties of their licensing environments does seem like an important first step toward a general understanding of the phenomenon.

4. *Alternative Approaches to the Sensitivity Problem*

The account developed here has three major virtues: first, by recognizing PSIs in general as a semantically coherent class of expressions, it explains their distributions directly in terms of the meanings they encode; second, the proposed classification provides a unified account for a wide range of both NPIs and PPIs; and finally, by distinguishing emphatic from understating PSIs, the account provides a principled explanation for distributional differences between two broad classes of PSI. But while the present proposal is more ambitious in scope than previous proposals, it has much in common with recent work on sensitivity by Krifka (1990, 1994), Kadmon and Landman (1993), Lee and Horn (1994), and more generally with the tradition following Ladusaw (1980) which views NPI licensing in terms of downward entailment. In this section I briefly consider alternative accounts of sensitivity. In Section 5 I will examine some differences between accounts based on downward entailment and the scalar model (SM) approach advocated here.

4.1. *Widening and Strengthening*

While the present paper offers a broad view of polarity sensitivity and PSIs in general, I have of necessity had little to say about individual PSIs.

By contrast, much work on polarity has been concerned almost exclusively with the English determiner *any*, and the semantic and syntactic complexities of this little word show that it more than deserves this attention. Recently, two new proposals have appeared offering unified accounts for the behavior of polarity sensitive (PS) and free choice (FC) *any* based directly on the word's meaning and pragmatic function. The proposal of Lee and Horn (henceforth L&H, 1994) mentioned above is quite congenial with the approach taken here: *any* is understood as an indefinite incorporating the semantics of *even*, and its peculiar distribution is explained as a consequence of its scalar semantics. Kadmon and Landman (henceforth K&L, 1993) offer an account which is in some ways very similar, contending that *any* is an indefinite determiner whose distribution is constrained by the interaction of two lexical semantic features, widening and strengthening. Unfortunately, a full discussion of either these proposals or even of *any* itself goes beyond the scope of this paper, but it may be useful to briefly consider the relative merits of K&L's analysis with respect to the scalar approach advocated here.

Traditionally, PS *any* (as in, "*I don't have any whiskey*") and FC any (as in, "*Mildred will drink any whiskey*") have been treated as homophones, the first viewed as having an existential, and the second, a universal force (Ladusaw, 1980; Carlson, 1980). Against this view, K&L argue that the dual patterning of *any* reflects its status as an indefinite determiner which, like the indefinite article *a*, can be interpreted either existentially or generically. This move allows K&L to sidestep the problem of the determiner's quantificational force by reducing the PS/FC contrast to an independently established pragmatic ambiguity. L&H offer further arguments, along with cross-linguistic evidence supporting this move, and in what follows I accept without comment the claim that NPs with PS *any* behave like regular indefinites while NPs with FC *any* behave like generic indefinites. My focus instead will be on the choice of widening and strengthening as the features to explain the indefinite's distribution.

K&L define widening and strengthening as follows:

(I)   WIDENING
        In an NP of the form *any CN*, *any* widens the interpretation of the common noun phrase (CN) along a contextual dimension. (1993, p. 369).

(II)   STRENGTHENING
        Any is licensed only if the widening that it induces creates a stronger statement, i.e., only if the statement on the wide

interpretation entails the statement on the narrow interpreta-
tion. (1993, p. 361).

For K&L, widening helps explain the fact that, as they put it, "*any*
indicates a reduced tolerance of exceptions" (1993, p. 356). Strengthening
functions as a constraint on the acceptability of a widened interpretation:
since the wide interpretation must entail the narrow, sentences like (42a)
are systematically ruled out while those like (42b) are systematically ac-
ceptable.

(42) a. *Mildred drank any whiskey.
     b.   Mildred didn't drink any whiskey.

(42a) is bad because the wide interpretation assertion that Mildred drank
any whiskey at all fails to entail the narrow reading that she drank some
particular whiskey; similarly, (42b) is good because the wide interpretation
that she didn't drink any whiskey at all does entail that she didn't drink
any particular whiskey. As in the scalar approach adopted here, K&L's
analysis successfully constrains the distribution of PS *any* in terms of the
interaction of two plausible semantic features. In some ways, in fact, the
two analyses seem to be notational variants: strengthening, for example,
amounts to much the same thing as saying that *any* encodes an emphatic
i-value such that the expressed proposition must entail an implicit cp.
Nonetheless, there are, I think, at least a couple of reasons to prefer a
scalar account in terms of q-value and i-value over the account in terms
of widening and strengthening. ·

First, there is the question of widening. K&L show that widening is a
feature of most sentences with *any* and they argue convincingly that this
feature arises independently from the emphatic stress that often falls on
any; however, there remain cases in which it is difficult to discern any real
widening, as in (43).

(43) a.   Does Mildred have any whiskey?
     b.   Mildred doesn't have any whiskey.

As K&L themselves acknowledge (1993, p. 363), examples such as these
sound rather neutral, a fact which suggests that widening might not be an
inherent characteristic of the word itself.

On the scalar model account, widening results from the increased pro-
minence which a PSI's quantitative value accords to a set of possible
alternative values. As argued in 3.1 above, a PSI's q-value refers to its
position within an ordered set of alternatives along some dimension of a
scalar model. For a form like *any*, which lacks any lexical content of its

own, this set of alternatives will be determined by the CN with which it combines, and in the default case the set will presumably exhaust the possible denotata of that CN. Thus, since expressions of the form *any* CN are scalar operators and must be interpreted with respect to an ordered set of alternative values, their use in context will necessarily highlight the full potential range of the head noun. And where this full range contrasts with some contextually constrained set of potential denotata, the effect of this highlighting will be precisely to widen the interpretation of the CN. If this is correct, then widening may not be an inherent part of the semantics of *any*, but rather a sort of pragmatic byproduct of its scalar semantics.[18] This would explain the apparent neutrality of examples like (43).

But the main advantage of the scalar account is its generality. K&L define both widening and strengthening as features peculiar to *any*. Thus while both are plausible lexical semantic features, they nonetheless lack independent motivation, and it is unclear how they might generalize to an account of other polarity items, particularly those which cannot be analyzed as indefinites. Of course, this in no way diminishes K&L's importance in providing a compelling and unified account of PS and FC *any*. My goal here is simply to suggest what might be gained by recasting their insights within a scalar account. By viewing *any* in particular and PSIs in general as scalar operators, the scalar account offers a way of extending K&L's insights about *any* to a unified account of polarity items in general, or at least to a significant portion of them. In any event, *any* will likely pose special problems for any theory of polarity.

4.2. *Polarity Lattices*

Krifka (1990, 1994) offers an account of polarity sensitivity in English and German that in many respects parallels the account developed here. Elaborating on the work of Heim (1984), Krifka proposes that PSIs are associated with a sort consisting of a quasi-ordered set of denotata in which the PSI is either the least element in the ordering (for negative polarity items) or the greatest element (for positive polarity items). As such, PSIs are understood as elements in a lattice structure. Krifka proposes a compositional mechanism which builds on this structure associating a complex expression containing a PSI (e.g. *I saw anything*) with a set of

---

[18] Of course widening might be conventionalized as a lexical feature in certain contexts or usages: FC *any*, for example, seems to require a widened or emphatic interpretation, while PS *any*, as (43) suggests, does not.

alternatives such that the complex expression is interpreted as either weaker (in the case of NPIs) or stronger (in the case of PPIs) than all its alternatives. Krifka then argues that PSIs are blocked in environments where they fail to be sufficiently informative. In other words, expressions containing PSIs must entail all other values in the PSI's polarity sort.

This, of course, is little more than a caricature: the importance, and the beauty, of an account such as Krifka's lies in the details of its formal workings. Nonetheless, it may be sufficient for a gross comparison with the scalar account proposed here. Both accounts view PSIs as invoking a set of alternatives ordered in terms of entailment relations, but there is an important difference in the way each handles the notion of informativeness. Crucially, Krifka does not recognize informative value as an inherent property of polarity items. Rather he derives the importance of informativity from the general Gricean principle that a speaker should say as much as is compatible with a given context. Furthermore, on Krifka's account, the problem with infelicitous uses of PSIs is not just that they are uninformative, but that they saliently evoke a set of more informative speech acts which are not performed. Thus an illicit PSI not only says too little, but it does so even as it evokes the possibility of saying more.

This account is somewhat suspect. As K&L argue "violation of Gricean maxims does not, in general, lead to ill-formedness – it does not render a sentence as hopelessly deviant as *I saw anything* is" (1993, p. 372). More importantly, even the most uninformative statement may serve a useful pragmatic purpose, and indeed, as argued above, many PSIs are in fact conventionally specified for just such a pragmatic purpose. Forms like *sorta*, *much* and *all that* all evoke a salient set of more informative propositions which could be but are not asserted. They thus appear only and precisely where they are maximally uninformative. By recognizing that i-value can be either high (hence emphatic and informative) or low (hence understating and uninformative) the SM account offers a natural explanation for this fact – a fact which Krifka seems to predict should not exist. Moreover, by positing informative value as a lexical semantic feature of PSIs (albeit a feature grounded in the pragmatics of informativeness), we avoid the pitfall of explaining grammatical well-formedness in terms of conversational principles: when a PSI is unacceptable it is not because the speaker could have said something better; rather it is because the PSI simply fails to express an essential part of its lexical semantics (K&L make a similar point with respect to strengthening, 1993, p. 373).

5. Licensing Alternatives: Downward Entailment and Negative Implicatures

Having discussed K&L's and Krifka's alternative views of the sensitivity problem, I turn now to the more general question of how PSIs are licensed. K&L and Krifka both take as their starting point Ladusaw's (1980, 1983) claim that NPIs are licensed in the scope of a downward entailing operator. The predictions of the scalar account concerning PSI licensing are not dramatically different from those made by a model like Ladusaw's, but the difference is, I suggest, significant.

Roughly, a downward entailing (DE) operator is one which allows subset for superset substitutions within its scope *salva veritate*. Other operators may be upward entailing (UE), allowing superset for subset substitutions *salva veritate*, or they may be nonmonotonic and so neither UE nor DE. More formally:

> A function f is downward entailing iff
>   for all X, Y in the domain of f, if $X \subseteq Y$, then $f(Y) \subseteq f(X)$.
> A function f is upward entailing iff
>   for all X, Y in the domain of f, if $X \subseteq Y$, then $f(X) \subseteq f(Y)$.

Many, probably most, of the common NPI licensers are in fact DE operators. The entailments in (36), for example, suggest that *rarely, few*, comparative clauses and relative clauses headed by a universally quantified NP all count as (or otherwise include) DE operators.

(44) a. Mookie rarely drinks milk. →
         Mookie rarely drinks skim milk.
   b. Few people understand the importance of syntactic theory. →
         Few people understand the importance of the minimalist program.
   c. Lou is too old to be spending all night at discos. →
         Lou is too old to be spending all night at Studio 54.
   d. Everyone who's eaten ice cream has had a taste of heaven. →
         Everyone who's eaten Vivoli's ice cream has had a taste of heaven.

In general, the environments picked out on the DE-account are pretty much the same as those picked out by the scale-reversal account; there are, however, significant theoretical and empirical differences between the two accounts. The fundamental theoretical difference between the two is that while Ladusaw defines licensing environments in terms of the truth-conditional semantics of scopal operators, the scalar model account defines

licensing environments in terms of the pragmatic interpretation of sentences in context.

In what follows, I will show that this pragmatic sensibility allows the SM approach to sidestep two major pitfalls of the logical DE account: it can handle licensing in environments which are not, strictly speaking, downward entailing, and it can handle the failure of licensing in environments which are incontrovertibly downward entailing. In other words, as suggested by Linebarger (1987, 1991) being in the scope of a DE operator turns out to be neither necessary nor sufficient for licensing PSIs; however, (pace Linebarger), licensing is based on a sort of inferencing – the pragmatic inferencing provided by an appropriately structured scalar model.

In considering the shortcomings of the DE account, it will be useful to keep Linebarger in mind. Linebarger (1980, 1987, 1991) proposes that NPIs are licensed primarily by occurrence in the immediate scope of negation at a syntactic level of logical form, and that NPIs may be licensed secondarily via a conventional negative implicature contributed by the NPI itself. According to Linebarger, in non-negative contexts, the NPI is licensed because it does occur in the immediate scope of negation in the logical form of the implicature. This proposal has come under attack for the rather unconstrained nature of a licensing mechanism based on implicature (Krifka, 1990; K&L, 1993; Yoshimura, 1994, among others). The worry is that without a precise account of how negative implicatures are generated, the theory will be immune from disconfirmation.[19] Moreover, the theory makes the questionable assumption that propositions based on conventional implicature are represented syntactically. These problems aside, Linebarger deserves credit for the insight that implicature may play a role in NPI licensing. I will suggest, however, that the role it plays is precisely to facilitate scalar inferencing.

5.1. *NPI Licensing without Downward Entailment*

5.1.1. *Exactly*

The first problems to consider are cases in which a clearly non-DE environ-

---

[19] Linebarger does in fact provide a number of interesting constraints on what can count as a negative implicature (NI). The **availability** requirement demands that a speaker be actively attempting to convey the NI. The **strength** requirement demands that the truth of the NI "must virtually guarantee" the truth of the overtly expressed proposition. And the **foreground** requirement demands that neither the NPI nor the NI occur as background information in the conversational context. I will not attempt to evaluate whether these constraints suffice to make her theory falsifiable.

ment does license negative polarity items. Linebarger (1991) discusses a variety of such cases, but I will confine myself to just two. The first of these, the use of *exactly* N as an NPI trigger, suggests that an account which allows a role for negative implicatures may succeed where an account based strictly on downward entailment fails.

Sentences with a subject modified by *exactly* pose a problem for the logical DE approach. *Exactly 3* is clearly neither upward nor downward entailing: from the truth of (45a) neither (45b) nor (45c) can safely be inferred.

(45) a. Exactly 3 professors read a novel last night.
 b. $\nrightarrow$ Exactly 3 professors read a book last night.
 c. $\nrightarrow$Exactly 3 professors read a trashy romance novel last night.

Exactly N is not UE since if just exactly three professors read a novel last night, it still may be that many more were busy reading important scholarly monographs. And *exactly* N is not DE since if exactly three professors read novels, it is still possible that they all read different kinds of novels: one might have read a trashy romance while the other two read trashy mysteries. These facts make the acceptability of a sentence like 46 problematic for a DE-based account.

(46)   Exactly three of the guests had so much as a drop of whiskey.

Linebarger suggests that such an example is acceptable because it can be taken as conveying the implicature that "most of the guests did not have so much as a drop of whiskey."

I think that Linebarger has the right intuition here. At first blush, the SM account seems to have the same problem with (46) that the DE account does: as the sentences in (47) suggest, *exactly* N is neither scale preserving nor scale reversing, any more than it is UE or DE.

(47) a. Exactly three professors can solve the hard puzzles.
       $\nrightarrow$ Exactly three professors can solve the easy puzzles.
 b. Exactly three professors can solve the easy puzzles.
       $\nrightarrow$ Exactly three professors can solve the hard puzzles.

In (47a), given three professors who can solve the hard problems, there is no reason to suppose that there isn't an abundance of professors who can solve the easy puzzles. In (47b), given only three who can solve the easy puzzles, it seems unduly optimistic to suppose that all of them could also solve the harder ones.

It should be clear that the reason (46) is well-formed has everything to do with the scalar semantics of *exactly*. The reason *exactly* can license

NPIs is that it adds something to what would be expressed by *3* N alone. While *3* N may sometimes be used to express "at least 3 N," *exactly 3* N makes explicit, and so indefeasible, the upper-bounding implicature "no more than 3 N." The sense of precision in a word like *exactly* is thus not symmetrical: although it means "neither more nor less" in practice the emphasis is often on the "not more" And in as much as *exactly 3* serves to convey *no more than 3* it is both downward entailing and scale-reversing: if at most 3 professors can solve the easy puzzles then at most 3 can solve the hard ones.

The suggestion here is that a sentence like (46) licenses NPIs not because of what it asserts or entails but more generally because of what it conveys. And crucially what it conveys in this case is not just a matter of truth conditional semantics, but also of the sentence's rhetorical function in context. Unless we allow the "no more than" reading as a conventional sense of *exactly*, a simple DE account cannot explain the licensing in (46): pragmatic considerations should not affect a monotonicity calculus. But they do affect scalar inferencing: given a party with a large number of guests, a sentence like (46) tells us not just how many people drank whiskey, but also how many, approximately, did not. And as Linebarger seems to suggest, the effect of the NPIs in (46) is precisely to emphasize the negative part of this conjunction.

In effect, what I am proposing is a compromise between Ladusaw and Linebarger. Linebarger is right to point to the importance of implicature to explain what licenses the NPI in (46); however, her account leaves the scalar nature of the implicature conveniently obscure. Ladusaw is right to point to the importance of inferencing as the crucial mechanism of licensing; however, his account leaves no room for pragmatic as opposed to logical inferencing. Ideally, we should have an account which could preserve the insights of both.

5.1.2. *Most*

The quantifier *most* provides our second example of an environment which at least sometimes licenses NPIs but which is nevertheless not downward entailing. As Ladusaw points out (1980, p. 151), *most* is difficult to judge, but, as the examples in (48) suggest, it seems to be neither upward nor downward entailing.

(48)a.   Most of the students who ate an apple got sick.
          ↛ Most of the students who ate some fruit got sick.

(48)b.   Most of the students who ate some fruit got sick.
         $\nrightarrow$ Most of the students who ate an apple got sick.

The inference in (48a) is invalid because it may be that while there were a lot of rotten apples, the rest of the fruit turned out to be fine. This shows that *most* is not UE on its first argument. The inference in (48b) is invalid because it may be that it was just those students who ate apples that avoided getting sick. This shows that *most* is not DE on its first argument.

Ladusaw welcomes these intuitions as they allow him to account for his judgement of (49) (his (53c)) as ill-formed.

(49) *Most of the students who had ever read anything about phren-
      ology attended the lecture.

I agree that (49) is less than beautiful, but it strikes me as somewhat less than horrible too. The reason for this, I think, is that while *most* is not strictly DE, it will nevertheless often allow outward inferencing in a scalar model. Thus when we apply the tests in (50), the result is slightly different from what we found in (48).

(50)a.   Most of the students who could solve the hardest puzzle got a
         prize.
         $\nrightarrow$ Most who could solve the easiest puzzle got a prize.
    b.   Most of the students who could solve the easiest problem got
         a prize.
         $? \rightarrow$ Most who could solve the hardest problem got a prize.

(50a) suggests that *most* does not license inward inferences on a scalar model: just because students who could solve the hardest puzzle were rewarded is no reason to assume that anything at all was given to students who performed the much less remarkable feat of solving the easiest puzzle. On the other hand, the outward, scale reversing inference in (50b) does seem to go through: if prizes are given for solving the easiest puzzle, then it seems natural to assume that prizes will also be given for solving the hardest puzzle. Given a scalar model pairing rewards on one dimension with accomplishments on another, general background assumptions will dictate that if small accomplishments (like solving an easy puzzle) merit certain rewards, then all greater accomplishments will merit rewards at least as great. Indeed, in order for the inference in (50b) to fail, we must either imagine a group of teachers who perversely value small achieve- ments over greater ones (such people do, of course, exist), or else assume that the awarding of prizes is only fortuitously related to the solving of

particular puzzles. Put slightly differently, the inference in (50b) will fail if and only if the sentence is not construed with respect to a scalar model.

Given these facts, the scalar model account of PSI licensing predicts that PPIs should be blocked in this context, but that NPIs should be acceptable so long as an appropriate scalar model is contextually available. These predictions gain support from the judgements in (51).

(51)a. ?Most of the students who studied an awfully long time got an
            A.
      b. ??Most of the students who studied at all wore earrings.
      c.   Most of the students who studied at all got an A.

The PPI *awfully* in (51a) is odd because its emphatic force would seem to suggest that the more students studied the less likely they were to get an A.[20] In (51b), the NPI *at all* is, at best, acceptable but bizarre because the scalar model required to license *at all* would have to somehow link studiousness with a proclivity for wearing earrings. Given normal background assumptions, this scalar model is simply not available. Finally, the same NPI in (51c) sounds perfectly normal because the required scalar model pairing studiousness with good grades does form a part of our stereotypical understanding of schoolwork.

While these facts clearly suggest an advantage of the scalar model account over the simple DE account, it is worth considering how Linebarger's negative implicature account might handle them. It is simply not at all obvious that a sentence like (51c) is associated with any negative implicature. The most likely candidate would perhaps be something like "If a student studied at all, then he got an A," which in turn would entail "either a student did not study at all or he got an A," and it would be this sentence which would license the NPI. Aside from the fact that this seems hopelessly convoluted, there are real problems both with the implicature and with the entailment. To begin with, one may question the appropriateness of imposing the logical equivalence of $(P \rightarrow Q)$ with $(\sim P \vee Q)$ on to the notoriously illogical English *if* and *or*. More impor-

---

[20] As pointed out by an anonymous reviewer, the same reasoning predicts that (i) should be fine. It's not, however,

(i)   ?Most of the students who studied an awfully long time got an F.

Note that the sentence does improve with *even* added before either *most* or *an awfully long time*, thus suggesting that an appropriate scalar model at least helps here. Apparently *awfully* has complexities of its own. My intuition is that it may be sensitive to information structure and require a context where it can function as new information. Or perhaps, as van der Wouden suggests for certain NPIs, it is barred from appearing in the restrictor of a quantificational trigger (1994, p. 73).

tantly, the gap between (51c) and the putative entailment from its putative implicature is just too significant. According to Linebarger's own constraints, the utterance of (51c) would have to commit the speaker to the truth of its licensing implicature and by extension all of its entailments. But this is clearly not the case as (52), which flatly contradicts the putative licenser, does not seem all that contradictory.

(52)   Most of the students who studied at all got an A, but doltish Dim only managed to get a B despite having studied for a good half an hour.

Thus while negative implicatures may be a normal concomitant of NPI usage, they do not appear to be necessary for licensing. The conclusion is that licensing within a scalar model is what really counts for a PSI's acceptability.

5.2. *Downward Entailment without NPI Licensing*

In this section I briefly consider cases in which NPIs are not licensed despite being in the scope of an appropriate downward entailing operator. Such cases show that the right sorts of entailment are of no avail if they cannot be construed as applying within a scalar model. I confine my attention here to two sets of examples, the first involving NPIs in relatives headed by a universal quantifier, the second involving NPIs in *before* clauses.

Although universal quantifiers are uncontroversially downward entailing on their first argument, they do not always manage to license NPIs. Linebarger (1980) and Heim (1984) discuss contrasts like those in (53)– (54).

(53)a.   Every restaurant that charges so much as a dime for iceberg lettuce ought to be closed down.
   b. ??Every restaurant that charges so much as a dime for iceberg lettuce actually has four stars in the handbook.
(54)a.   Anyone who gives a damn about the environment will enjoy recycling.
   b. ??Anyone who gives a damn about the environment will shop at the Gap.

As Heim notes, the intuitive difference between the a- and b-sentences in these examples is that in the a-sentences, but not the b-sentences, there is some natural connection between the matrix and relative clauses. As she puts it,

the predicate in 53a (her (36)) is something that applies to restaurants *because* they charge
a dime or more for iceberg lettuce . . . whereas the predicate in (53b) (her (37)) just *happens*
to apply to those restaurants without regard to, or even in spite of, what they charge for
iceberg lettuce. (1984, pp. 104–5).

Heim then goes on to argue that the reason the b-sentences aren't accept-
able is that the NPIs in them somehow incorporate the semantics of *even*.
This means that a sentence like (53b) will imply that "that there are values
other than one dime for x which make 'Every restaurant that charges x
for iceberg lettuce ought to be shut down' true" (ibid, p. 106). But of
course this condition is just a special case of the scalar model account:
since NPIs like *a red cent* and *so much as* are emphatic scalar operators,
their interpretation is analogous to that of the non-PSI emphatic scalar
operator *even*. Because all these forms are scalar operators, they require
the availability of some dimension of values that can contrast with the
value picked out by the scalar operator. The b-sentences are thus bad
because they don't allow for the easy construal of a cp that can make the
tp emphatic.

Yoshimura (1994) notes a similar phenomenon with respect to NPIs in
before clauses.

(55) a. Miss. Prism spilled her wine before she had drunk a drop.
     b. ??Miss. Prism poured her wine before she had drunk a drop.
(56) a. The alarm clock was ringing before I managed to sleep a wink.
     b. ??It was raining before I managed to sleep a wink.
(57) a. Oscar had been studying linguistics for 10 years before he
         learned a damned thing about pragmatics.
     b. ??Oscar had been fishing many times before he learned a damn
         thing about pragmatics.

Once again, NPIs in the a-sentences are acceptable because they express
minimal degrees which would naturally be expected to obtain before the
reference time marked in the matrix clause. In the b-sentences, where
normal assumptions about the world will not supply an appropriate con-
nection between the NPI and the passage of the reference time, the NPI
is at best rather peculiar sounding. (55) is particularly instructive in this
regard since there is a natural connection between pouring wine and
drinking it: as a rule, given normal social conventions and barring the use
of straws, until something is poured, it cannot be drunk. But this entail-
ment is absolute: waiting longer before pouring something does not, under
normal conditions, increase the likelihood that any quantity will be drunk;
however, waiting longer before spilling something may well have this

effect. For this reason, the a-sentence is easily construed within a scalar model, while the b-sentence is not.

Comparable examples may be adduced with just about any potential trigger. (58) illustrates this with *without* and (59) shows that even sentential negation cannot make sense of a minimizer if it cannot be construed within a scalar model.

(58) a. Algernon left without saying a word.
    b. ??Algernon enjoyed the movie without saying a word.
(59) a. Cecily didn't eat a bite of her food.
    b. ??Cecily didn't stare at a bite of her food.

(58a) is appropriately emphatic since in the normal case one is supposed to say something before taking one's leave; on the other hand, since enjoying a movie normally is not correlated with the amount one talks, the minimizer *a word* seems oddly out of place. Similarly, for (59), there are many activities for which a bite of food might count as a natural minimal unit, but staring is not one of them: one can just as easily stare at a whole banquet as at a single bite, and so the minimizer in this context fails to supply a particularly emphatic proposition.

It is not enough simply to be downward entailing: it also matters just what is being entailed. Where an appropriate scalar model is available, NPIs are licensed. Where no scalar model can be supplied, NPIs are unwelcome.

6. CONCLUSIONS

This paper began with an effort to restate the problem of polarity sensitivity in the broadest terms possible, distinguishing three major issues – licensing, sensitivity and diversity – which any complete theory should address. In my analysis, though I have not presented a complete theory myself, I have made some rather ambitious claims. I have sought a unified and comprehensive account of polarity sensitivity, arguing that PSIs in general are scalar operators and that two independently motivated lexical semantic features, quantitative and informative value, account for the peculiar distributions that define the phenomenon. The analysis receives considerable support from a large range of polarity items, in English and other languages, which transparently encode a scalar semantics and which divide rather neatly into four major classes based on the interaction of q-value and i-value. Furthermore, I have argued that a pragmatic, scalar model account allows for important refinements in explaining where and why PSIs are licensed.

Still, one may reasonably question whether this approach really can extend to all polarity items, or even whether we should expect a general unified account of polarity sensitivity. Indeed, the conventional wisdom in some circles is that "there is no universal explanation for the existence of all polarity items" (van der Wouden, 1994, p. 91) and that only by recognizing a variety of factors which draw lexical items to negative contexts will we ever gain insight into the diversity of polarity phenomena (cf. Hoeksema, 1994; Rullmann, 1996). Viewed in this light, the present paper would seem to err in its quest for universality by underestimating or ignoring the complexity of the data. Certainly, many mysteries remain. More work is needed to understand how and why different contexts facilitate different kinds of scalar inferencing. More is needed to understand what factors determine the availability of an appropriate scalar model. And much more is needed to understand the many subtle differences in form, function and sensitivity between different polarity items.

The goal, however, has not been to solve every mystery, but rather just to find a general framework in which they might all be related. The claim is not that all polarity items are in any way equivalent, only that at a very schematic level of representation they all share certain basic semantic features. Different scalar operators, and different PSIs, can and do differ dramatically – in terms of the sorts of scales they evoke, their grammatical function, semantic and formal complexity, degree of grammaticization and conventionalization, and frequency and register of use, to name just a few parameters. The modest claim of this paper is just that beneath this teeming diversity lies an essential unity, and that this unity is a matter of lexical semantics.

Ultimately, whether or not all PSIs do fit into the scalar approach, it should be clear that scalar semantics, and in particular the basic scalar features of q-value and i-value do play a crucial role in the sensitivities of a great many polarity items. This is significant because, among other things, it suggests that polarity sensitivity need not be considered an arbitrary grammatical phenomenon. As such, the analysis presented here conforms to the content requirement for Cognitive Grammar, which prohibits arbitrary formal devices in linguistic explanations (Langacker, 1987, p. 53). Polarity sensitivity need not be stipulated as a distributional constraint; rather it simply reflects a particular encoding of basic lexical semantic features.

These lexical features, i-value and q-value, are hardly arbitrary formal devices. Scalar reasoning would seem to be a clear example of a basic and universal human cognitive ability, one which is clearly not specific to the domain of language. Indeed, considering their remarkable schematicity

and pervasiveness throughout the lexicon, i-value and q-value and, more generally, the ability to reason within a scalar model, seem like natural candidates for conceptual primitives. If this is correct, then it is surely significant, for it offers a clear example of how conventional grammatical phenomena may be driven by general cognitive capacities.

Finally, and perhaps most importantly, the analysis presented here allows us to clear up, or at least shed some light on, what may be the most vexing question about polarity sensitivity: namely, why should polarity items exist at all? What good does it do for a language to have forms which cannot appear in certain contexts? Why should such a patently dysfunctional phenomenon be so pervasive in the languages of the world? On the present account, polarity items are really not peculiar at all: as with any lexical form, their distributions are constrained by the meanings they encode. Polarity, of course, remains a grammatical phenomenon, for whether or not a given form is conventionally specified for i-value and q-value is an arbitrary linguistic fact. But in the end, polarity items exist because they are useful. Indeed, their pervasiveness is a testimony to the fundamental importance of scalarity and informativity as basic aspects of human cognition and communication.

REFERENCES

Anscombre, Jean-Claude and Oswald Ducrot: 1983, *L'Argumentation dans la Langue*, Mardaga, Bruxelles.
Baker, C. L.: 1970, 'Double Negatives', *Linguistic Inquiry* 1, 169–86.
Bolinger, Dwight: 1972, *Degree Words*, Mouton, The Hague.
Borkin, Ann: 1971, 'Polarity Items in Questions', *CLS* 7, 53–62.
Brown, Penelope and Stephen Levinson: 1978, 'Universals in Language Usage: Politeness Phenomena', in E. Goody (ed.), *Questions and politeness*, Cambridge University Press, Cambridge, pp. 56–289.
Carlson, Greg: 1980, 'Polarity *Any* is Existential', *Linguistic Inquiry* 11, 799–804.
Edmondson, J. A.: 1981, 'Affectivity and Gradient Scope', *CLS* 17, 38–44.
Edmondson, J. A.: 1983, 'Polarized Auxiliaries', in F. Heny and B. Richards (eds.), *Linguistic Categories: Auxiliaries and Related Puzzles*, Vol. I, D. Reidel Publishing Company, Dordrecht, pp. 49–68.
Fauconnier, Gilles: 1975a, 'Polarity and the Scale Principle', *CLS* 11, 188–99.

Fauconnier, Gilles: 1975b, 'Pragmatic Scales and Logical Structures', *Linguistic Inquiry* **6**, 353–75.

Fauconnier, Gilles: 1979, 'Implication Reversal in a Natural Language', in F. Guenther and S. J. Schmidt (eds.), *Formal Semantics and Pragmatics for Natural Languages*, D. Reidel, Dordrecht, pp. 289–301.

Fauconnier, Gilles: 1980, *Etude de Certains Aspects Logiques et Grammaticaux et de la Quantification et de L'Anaphore en Français et en Anglais*, Thèse présentée devant L'Université de Paris VII, 1976. Atelier Reproduction des Thèses, Lille.

Fillmore, Charles J., Paul Kay, and Mary Catherine O'Connor: 1988, 'Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*', *Language* **64**, 501–38.

Francescotti, Robert M.: 1995, 'Even: the Conventional Implicature Approach Reconsidered', *Linguistics and Philosophy* **18**, 153–173.

Giannakidou, Anastasia: 1994, 'The Semantic Licensing of Negative Polarity Items and the Modern Greek Subjunctive', in *Language and Cognition 4: Yearbook of the Research Group for Theoretical and Applied Linguistics of the University of Groningen*.

Haspelmath, Martin: 1993, 'A Typological Study of Indefinite Pronouns', Ph.D. thesis, Freie Universität Berlin.

Heim, Irene: 1984, 'A Note on Negative Polarity and Downward Entailingness', *Proceedings of NELS* **14**, 98–107.

Hinds, Marilyn: 1974, 'Doubleplusgood Polarity Items', *CLS* **10**, 259–268.

Hoeksema, Jack: 1983, 'Negative Polarity and the Comparative', *Natural Language and Linguistic Theory* **1**, 403–434.

Hoeksema, Jack: 1994, 'On the Grammaticalization of Negative Polarity Items', *Proceedings of the Berkeley Linguistics Society* XX.

Horn, Laurence R.: 1969, 'A Presuppositional Analysis of *only* and *even*', *CLS* **5**, 97–108.

Horn, Laurence R.: 1970, 'Ain't it Hard (anymore)', *CLS* **6**, 318–327.

Horn, Laurence R.: 1972, 'On the Semantic Properties of Logical Operators in English', Ph.D. dissertation, distributed by IULC, 1976.

Horn, Laurence R.: 1985, 'Metalinguistic Negation and Pragmatic Ambiguity', *Language* **61**, 121–74.

Horn, Laurence R.: 1989, *A Natural History of Negation*, University of Chicago Press, Chicago and London.

Horn, Laurence R.: 1996, 'Exclusive Company: *Only* and the Dynamics of Vertical Inference', *Journal of Semantics* **13**, 1–40.

Hübler, Axel: 1983, *Understatements and Hedges in English*, John Benjamins, Amsterdam/ Philadelphia.

Israel, Michael: 1995a, 'Negative Polarity and Phantom Reference', *Proceedings of BLS* **XXI**, 162–173.

Israel, Michael: 1995b, 'Review of Negative Contexts by Ton van der Wouden', *Glot International* **1**(5), 10–12.

Israel, Michael: 1995c, 'The Scalar Model of Polarity Sensitivity', paper delivered at the *Ottawa Conference on Negation: Syntax and Semantics*.

Kadmon, Nirit and Fred Landman: 1993, 'Any', *Linguistics and Philosophy* **16**, 353–422.

Kay, Paul: 1983, 'Linguistic Competence and Folk Theories of Language: Two English Hedges', *Proceedings of BLS* **IX**, 128–37.

Kay, Paul: 1989, 'Contextual Operators: *respective*, *respectively*, and *vice versa*', *Proceedings of BLS* **XV**, 181–192.

Kay, Paul: 1990, 'Even', *Linguistics and Philosophy* **13**, 59–111.

Klima, Edward S.: 1964, 'Negation in English', in J. Fodor and J. Katz (eds.), *The Structure of Language: Readings in the Philosophy of Language*, Prentice-Hall, Englewood Cliffs, N.J., pp. 246–323.

Krifka, Manfred: 1990, 'Some Remarks on Polarity Items', in D. Zaefferer (ed.), *Semantic Universals and Universal Semantics*, Foris, Dordrecht, pp. 150–189.

Krifka, Manfred: 1994, 'The Semantics and Pragmatics of Weak and Strong Polarity Items in Assertions', *Proceedings of SALT IV*, 195–219.

Ladusaw, William: 1980, *Polarity Sensitivity as Inherent Scope Relations*, Garland Publishing, New York and London.

Ladusaw, William: 1983, 'Logical Form and Conditions on Grammaticality', *Linguistics and Philosophy* **6**, 373–92.

Laka, Itziar: 1990, 'Negation in Syntax: On the Nature of Functional Categories and Projections', Unpublished Ph.D. Dissertation, MIT.

Lakoff, George: 1972, 'Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts', *CLS* **8**, 183–228.

Lakoff, Robin: 1973, 'The Logic of Politeness; or, minding your p's and q's', *CLS* **9** 149–62.

Langacker, Ronald W.: 1987, *Foundations of Cognitive Grammar vol. 1: Theoretical Prerequisities*, Stanford University Press, Stanford.

Lee, Young-Suk and Laurence Horn: 1994, '*Any* as Indefinite plus *Even*', Ms. Yale University.

Linebarger, Marcia: 1980, 'The Grammar of Negative Polarity', Ph.D. Dissertation, MIT.

Linebarger, Marcia: 1987, 'Negative Polarity and Grammatical Representation', *Linguistics and Philosophy* **10**, 325–87.

Linebarger, Marcia: 1991, 'Negative Polarity as Linguistic Evidence', *Papers from the Parasession on Negation, CLS* **27**, 165–188.

Löbner, Sebastian: 1987, 'Quantification as a Major Module of Natural Language Semantics', in J. Groenendijk et al. (eds.) *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, pp. 53–85. Foris: Dordrecht.

Löbner, Sebastian: 1989, 'German *schon-erst-noch*: An Integrated Analysis', *Linguistics and Philosophy* **12**, 167–212.

Michaelis, Laura A.: 1992, 'Aspect and the Semantics-pragmatics Interface: The Case of *already*', *Lingua* **87**, 321–339.

Michaelis, Laura A.: 1993. '"Continuity" within Three Scalar Models: The Polysemy of Adverbial *still*', *Journal of Semantics* **10**, 193–237.

Moxey, Linda M. and Anthony J. Sanford: 1993, *Communicating Quantities: A Psychological Perspective*. Hove (UK) and Hillsdale (USA): Lawrence Erlbaum Associates, Ltd.

Moxey, Linda M. and Anthony J. Stanford: 1994, 'Psychological Studies of Quantifiers', *Journal of Semantics* **11**(3), 153–170.

Progovac, Ljiljana: 1992, 'Negative Polarity: A Semantico-Syntactic Approach', *Lingua* **86**, 271–299.

Progovac, Ljiljana: 1994, *Negative and Positive Polarity*, Cambridge University Press, Cambridge.

Raghibdoust, Shahla: 1994, 'The Semantic-Pragmatic Nature of the Persian Polarity Items', Ms. University of Ottawa.

Rullmann, Hotze: 1996, 'Two Types of Negative Polarity Items', *Proceedings of NELS* 26.

Schmerling, Susan: 1971, 'A Note on Negative Polarity', *Papers in Linguistics* **4**, 200–206.

Spitzbardt, Harry: 1963, 'Overstatement and Understatement in British and American English', *Philologica Pragensia* **6**(45), 277–286.

Stern, Gustaf: 1931, *Meaning and Change of Meaning*, Indiana University Press, Bloomington.

Sweetser, Eve: 1990, *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*, Cambridge University Press, Cambridge.

Traugott, Elizabeth Closs: 1988, 'Pragmatic Strengthening and Grammaticalization', *Proceedings of BLS* XIV, 204–214.

Traugott, Elizabeth Closs and Ekkehard König: 1991, 'The Semantics-pragmatics of Grammaticalization Revisited', in E. C. Traugott and B. Heine (eds.), *Approaches to Grammaticalization*, vol. 1, John Benjamins, Amsterdam/Philadelphia, pp. 189–218.

Uribe-Etxebarria, María: 1994, *Interface Licensing Conditions on Negative Polarity Items: A Theory of Polarity and Tense Interactions*, Ph.D. Dissertation, University of Connecticut.

van der Auwera, Johan: 1993, '*Already* and *Still*: Beyond Duality', *Linguistics and Philosophy* **16**, 613–53.

van der Wouden, Ton: 1994, *Negative Contexts,* Groningen Dissertations in Linguistics: University of Groningen.

van Os, Charles: 1989, *Aspekte der Intensivierung im Deutschen*, Tübingen: Gunter Narr Verlag.

Verhagen, Arie: 1995, 'Meaning and the Coordination of Cognition', Paper presented at ICLA 4, Albuquerque, New Mexico.

von Bergen, Anke and Karl von Bergen: 1993, *Negative Polarität im Englischen*, Narr Tübingen.

Yoshimura, Akiko: 1994, 'A Cognitive Constraint on Negative Polarity Phenomena', *Proceeedings of BLS XX.*

Zwarts, Frans: 1990, 'The Syntax and Semantics of Negative Polarity', to appear in *Views on the Syntax-Semantics Interface II*, S. Busemann et al. (eds.), Berlin.

University of California – San Diego
Department of Linguistics 0108
La Jolla, CA 92093-0108
U.S.A.