87. Scalar implicature as a grammatical phenomenon

Abstract

This article develops various arguments for the view that scalar implicatures should be derived within grammar and not by a theory of language use (pragmatics). We focus primarily on arguments that scalar implicatures can be computed in embedded positions, a conclusion incompatible with existing pragmatic accounts. We also briefly review additional observations that come from a variety of empirical domains, all incompatible with pragmatic accounts, yet predicted by the grammatical alternative.

1. Introduction

Since the late 1990's there has been a lively revival of interest in implicatures, particularly scalar implicatures (SIs for short). Building on the resulting literature, our main goal in the present article is to present several arguments for the claim that SIs can occur systematically and freely in arbitrarily embedded positions. We are not so much concerned with the question of whether drawing implicatures is a costly option (in terms of semantic processing, or of some other markedness measure). Nor are we specifically concerned with how implicatures come about (even though, to get going, we will have to make some specific assumptions on this matter). The focus of our discussion is testing the claim of the pervasive embeddability of SIs in just about any context, a claim that remains so far controversial. While our main goal is the establishment of an empirical generalization, if we succeed, a predominant view on the division of labor between semantics and pragmatics will have to be revised. A secondary goal of this article is to hint at evidence that a revision is needed on independent grounds. But let us first present, in a rather impressionistic way, the reasons why a revision would be required if our main generalization on embedded SIs turns out to be correct.

In the tradition stemming from Grice (1989), implicatures are considered a wholly pragmatic phenomenon (cf. article 92 (Simons) *Implicature*, article 5 (Green) *Meaning in language use*, and also Davis 2005) and SIs are often used as paramount examples. Within such a tradition, semantics is taken to deal with the compositional construction of sentence meaning (a term which we are using for now in a loose, non technical way), while pragmatics deals with how sentence meaning is actually put to use (i.e. enriched and possibly modified through reasoning about speakers' intentions, contextually relevant information, etc.). Simply put, on this view pragmatics takes place at the level of complete utterances and pragmatic enrichments are a root phenomenon (something that happens globally to sentences) rather than a compositional one. So if SIs can be systematically generated in embedded contexts, something in this view has got to go. Minimally, one is

Maienborn, von Heusinger and Portner (eds.) 2012, Semantics (HSK 33.3), de Gruyter, 2297–2331

forced to conclude that SIs are computed compositionally on a par with other aspects of sentence meaning. But more radical task reallocations are also conceivable. While we may not be able to reach firm conclusions on this score, we think it is important to arrive at a consensus on what are the factual generalizations at stake, how they can be established, and what range of consequences they may have.

Let us rephrase our point more precisely. The semantics/pragmatics divide can usefully be lined up with compositional vs. postcompositional interpretive processes. In the compositional part, basic meanings are assigned to lexical entries, which are then composed bottom up using a restricted range of semantic operations on the basis of how lexical entries are put together into phrases. These operations apply in an automatic fashion, blind to external considerations, e.g., speaker intentions and relevant contextual knowledge. Sentence meaning is, thus, constructed through the recursive application of semantic rules – typically, functional application. But what is sentence meaning? Such a notion is often identified with truth conditions. While semantics, as we understand it, falls within this tradition, we would like to keep our options open on the exact nature of sentence meaning. For the notions of sentence content that have emerged from much recent work are way more elaborate than plain truth conditions (cf. article 71 (Hinterwimmer) *Information structure* and 94 (Potts) *Conventional implicature*). For example, sentence meaning has been argued to involve the computation of alternative meanings and hence to be a multidimensional phenomenon (cf. the semantics of questions, focus, etc); or sometimes sentence meaning has been assimilated to context change potentials (cf. dynamic approaches to presuppositions and anaphora). We remain neutral here on these various options, and we do so by simply taking sentence meaning as equivalent to the output of the compositional process of interpretation as determined by UG, whatever that turns out to be.

In understanding the compositional/postcompositional divide, one further preliminary *caveat* must be underscored. Sentence meaning is blind to context, but not independent of it. Virtually every word or phrase in Natural Language is dependent on the context in some way or other. In particular, the meaning of sentences will contain variables and indexicals whose actual denotation will require access to factual information accessible through the context. To illustrate, consider the following standard analysis of *only* and focus association, along the lines of Rooth (1985, 1992) and Krifka (1993) (an example that will turn out to be very useful for our approach to SIs). According to Rooth a sentence like (1a) is analyzed along lines explicated in (1b-d):

(1) a. Joe only upset [$_F$Paul and Sue]
(where [$_F$] indicates the constituent bearing focal stress)
b. **LF:** only [Joe upset [$_F$Paul and Sue]]
c. **Interpretation:**
$[[\text{Only}_{\text{ALT(D)}}\text{Joe upset } [_F\text{Paul and Sue}]]]^{w_0} = 1$ iff
$\text{UPSET}(\text{JOHN}, \text{PAUL} + \text{SUE})\ (w_0) = 1 \wedge \forall p \in \text{ALT(D)}\ [\lambda w.\text{UPSET}(\text{JOHN}, \text{PAUL} + \text{SUE})\ (w) \not\subseteq p \rightarrow p(w_0) = 0]$
d. $\text{ALT(D)} = \{\lambda w.\text{UPSET}(\text{JOHN}, u)\ (w) : u \in D\} =$
{The proposition that Joe upset Lee, the proposition that Joe upset Sue, the proposition that Joe upset Kim, the proposition that Joe upset Lee and Sue, . . . }

Something like (1a) has the Logical Form in (1b), where *only* is construed as a sentential operator, and is interpreted as in (1c). Such an interpretation, informally stated, says that

Joe upset Paul and Sue and that every member of the contextually restricted set of alternatives ALT not entailed by the assertion must be false. Thus, in particular, *Joe upset Paul* is entailed by the assertion, and hence has to be true, but *Joe upset Kim* is not, and hence must be false. The set ALT is specified as in (1d). Such a set is generated by UG-driven principles through a separate recursive computation (and this is part of what makes sentence meaning multidimensional). In (1c), there is one variable whose value has to be picked up by pragmatic means: D, the quantificational domain. The determination of D's value is a pragmatic, 'postcompositional' process.

So, pragmatics, as understood here, is the process whereby speakers converge on reasonable candidates as to what the quantificational domain may be; it is also the process whereby a sentence like (1a) may wind up conveying that the meeting was a success (because, say, Joe managed to keep the number of upset people to a minimum), or the process whereby (1a) may result in an ironical comment on Joe's diplomatic skills, etc. Such processes are arguably postcompositional, in the sense that they presuppose a grasp of sentence meaning, plus an understanding of the speaker's intentions, etc. We have no doubt that such processes exist (and, thus, that aspects of the Gricean picture are sound and effective). The question is whether SIs are phenomena of the latter postcompositional sort or are UG-driven like, say, the principles of focus association sketched in (1).

1.1. Background

In his seminal work, Grice (1989) argues that the main source of pragmatic enrichment is a small set of maxims (Quality, Quantity, Relation, Manner) that govern, as overridable defaults, cooperative conversational exchanges (Oswald Ducrot developed related ideas independently, e.g. Ducrot 1973. Another approach, broadly inspired by Grice but which departs more radically from the original formulations, can be found within the tradition of Relevance Theory – see, e.g., Sperber & Wilson 1986, Carston 1988).

In discussing the various ways in which these maxims may be used to enrich basic meanings, Grice considers the case of how *or* might strengthen its classical Boolean inclusive value (according to which 'p or q' is true if at least one of the two disjuncts is true) to its exclusive construal ('p or q' is true if one and only one of the two disjuncts is). In what follows, we offer a reconstruction of the relevant steps of this enrichment process, as is commonly found in the literature (cf., e.g., Gamut 1991). The basic idea is that, upon hearing something like (3a), a hearer considers the alternative in (3b) and subconsciously goes through the reasoning steps in (3i-vi)

(3) a. Joe or Bill will show up
b. Joe and Bill will show up
i. The speaker said (3a) and not (3b), which, presumably, would have been also relevant [Relevance]
ii. (3b) asymmetrically entails (3a), hence is more informative
iii. If the speaker believed (3b), she would have said so [Quantity]
iv. It is not the case that the speaker believes that (3b) holds
v. It is likely that the speaker has an opinion as to whether (3b) holds.
Therefore:
vi. It is likely that the speaker takes (3b) to be false.

This example illustrates how one might go from (3a) to (3vi) by using Grice's maxims and logic alone. The conclusion in (3vi) is close to the desired implicature but not quite. What we actually want to draw is that the speaker is positively trying to convey that Joe and Bill will not both come. Moreover, we need to be a bit more precise about the role of relevance throughout this reasoning, for that is a rather sticky point. We will do this in turn in the next three subsections.

1.2. SIs as exhaustifications

To understand in what sense the conclusion in (3vi) should and could be strengthened, it is convenient to note that the reasoning in (3) can be viewed as a form of *exhaustification* of the assertion, i.e., tantamount to inserting a silent *only*. Using B_s as a short form for 'the speaker believes that', the assertion in (3a) would convey to the reader the information in (4a), while the alternative assertion in (3b) would convey (4b).

(4) a. B_s (show up(j) $\vee$ show up(b))
b. B_s (show up(j) $\wedge$ show up(b))

If you now imagine adding a silent *only* (henceforth, *O*) to (4a) (and evaluating it with respect to the alternative in (4b)), we get:

(5) $O_{ALT}(B_s$ (show up(j) $\vee$ show up(b)))
= B_s (show up(j) $\vee$ show up(b)) $\wedge\neg$ B_s (show up(j) $\wedge$ show up(b))

The result in (5) is the same as (3iv) and entitles the hearer only to the weak conclusion in (3vi) (and for the time being, we might view this use of *O* as a compact way of expressing the reasoning in (3)). Now, the conclusion we would want instead is:

(6) B_s (O_{ALT}(show up(j) $\vee$ show up(b)))
= B_s (show up(j) $\vee$ show up(b) $\wedge$ $\neg$ (show up(j) $\wedge$ show up(b)))

The speaker, in other words, by uttering (3a), is taken to commit herself to the negation of (3b). The reading in (6) can be derived if we are allowed to go from something like *it is not the case that x believes that p* to *x believes that not p*. Sauerland (2004) calls this 'the epistemic step'. What is relevant in the present connection is that in the computation of SIs such a step does not follow from Gricean maxims and logic alone. It is something that needs to be stipulated. This seems to be a gap in the Gricean account of SIs (see for instance Soames 1982, Groenendijk & Stokhof 1984). And this problem interacts with another, even more serious one, having to do with seemingly innocent assumption that in uttering (3a), something like (3b) is likely to be relevant. Let us discuss it briefly.

1.3. Relevance

Let us grant that in uttering (3a), (3b) is also indeed relevant, whatever 'relevant' may mean. Now, a natural assumption is that the property of 'being relevant' is closed under negation, i.e., if a proposition ϕ is relevant, then $\neg\,\phi$ is relevant as well. To say that ϕ is relevant must

be to say that it matters whether ϕ is true or false (this follows from several formal definitions of relevance proposed in the literature, e.g., Carnap 1950; Groenendijk & Stokhof 1984, 1990). If this is so, the negation of (3b) will also be relevant. But then the set of relevant alternatives changes. Assuming that relevance is also closed under conjunction (if A and B are both relevant, then so is *A and B*), it now includes:

(7) a. show up(j) ∨ show up(b)
b. show up(j) ∧ show up(b)
c. ¬(show up(j) ∧ show up(b)))
d. (show up(j) ∨ show up(b)) ∧ ¬(show up(j) ∧ show up(b)))

Now, note that both (7d) and (7b) a-symmetrically entails (7a) (i.e., (3a)). So if we run the Gricean reasoning in (3) over this expanded set of alternatives or, equivalently, if we exhaustify the assertion along the lines discussed in (4), we must conclude that the speaker's only relevant belief is (7a), and, in particular, that he does not have the belief that (7b) is true, nor than (7d) is true. In other words, we conclude that a) the speaker believes that John or Bill will show up, b) that she does not have the belief that both will show up and c) that she does not have the belief that only one of the two will show up. Notice that in this case, the epistemic step would lead to a contradiction. In other words, we have to conclude that if the speaker utters *p or q*, he must not have an opinion as to whether *p and q* is the case, which blocks the possibility that he believes *p and q* to be false (a precursor of this argument, which was made explicit by von Fintel & Heim in their 1997 pragmatics class notes, can be found in Kroch 1972 – see also Davis 1998). This problem is a general one: the assumption that relevance is closed under negation (which is hard to avoid) has the effect of blocking any potential SI.

So we see that on the one hand, by logic alone, we are not able to derive SIs in their full strengths from the Gricean maxims. And, if we are minimally explicit about relevance, we are able to derive no implicature at all (except for 'ignorance' implicatures). Something seems to be going very wrong in our attempt to follow Grice's ideas. However, post-Gricean scholars, and in particular Horn (1972, 1989), have addressed some of these problems and it is important to grasp the reach of such proposals.

1.4. Scales

Horn's important point is that if we want to make headway in understanding how SIs come about, then the set of relevant alternatives needs to be constrained. In the most typical cases, they will be *lexically* constrained by items of the same category whose entailments line them up in a scale of increasing informativeness. Examples of Horn's scales are the following:

(8) a. The positive quantifiers: *some*, *many*, *most*, *all*
b. The negative quantifiers: *not all*, *few*, *none*
c. Numerals: *one*, *two*, *three*, ...
d. Modals: *can*, *must*
e. Sentential connectives: *or*, *and*
f. Gradable adjectives: *warm*, *hot*, *boiling* / *chilly*, *cold*, *freezing*, etc.

These series are characterized by the fact that the items on the right are stronger than the items on their left. For example, if all of the students did well, then most of them did and surely some of them did. Similarly for the other scales. Horn's proposal is that if you use *some*, other members of the scale may be activated and provide the alternatives against which the assertion is evaluated. Not all have to be activated; perhaps none of them will. But if they are activated, they must look like in (8). What is crucial about these scales is that one cannot mix elements with different monotonicity/polarity properties (see Fauconnier 1975b and Matsumoto 1995). Thus for example, one cannot have positive and negative quantifiers as part of the same scale. This is the way the problem considered in section 1.3. is circumvented.

Horn's suggestions can be extended to other seemingly more volatile/ephemeral scales. Consider the following example, modeled after Hirschberg (1985):

(9) A: Did John mail his check?
B: He wrote it.

This dialogue suggests that B's intention is to convey that John didn't mail the check. The 'scale' being considered here must be something like {write the check, mail the check}. What is crucial is that we do not consider mailing vs. not mailing, or mailing vs. stealing, for otherwise we would only derive ignorance implicatures (on the role played by questions for determining alternatives, see also van Kuppevelt 1996 and Spector 2006).

The main moral is that the notion of 'relevance' to be used in implicature calculation is, yes, context dependent but constrained in at least two ways: through the lexicon (certain classes of words form lexical scales) and through a monotonicity constraint: all scales, even scales that are not lexically specified, such as those needed for (9), cannot simultaneously include upward and downward entailing elements.

1.5. Monotonicity and scale reversal

There is a further important point to make. Let us consider an example like (10).

(10) A: Who will come to the party?
B: I doubt that Joe or Sue will come.

Here no implicature comes about, even though the conjunctive statement *Joe and Sue will come to the party* must be relevant on the background of the question in A. The reason is the following. The active alternative to B is *I doubt that Joe and Sue will come to the party*. Since B's utterance *entails* this alternative, the latter cannot be excluded and no implicature comes about (operators which, like negation and the verb *doubt*, reverse the direction of entailment are called downward entailing, or monotone decreasing; operators which *preserve* entailment patterns are called upward entailing, or monotone increasing). In our terms, use of covert *only* in cases like these is simply vacuous. It is useful to compare (10) with:

(11) A: Who will come to the party?
B: I doubt that all of the students will
B-ALT: I doubt that (some of the) students will come to the party

If B's answer is as indicated, its alternative would presumably be something like B-ALT. (B-ALT might be somewhat less than felicitous because *some* is a positive polarity item – whence the parentheses.) So, exhaustifying B's utterance will bring about the negation of B-ALT, namely:

(12) It is not true that I doubt that (some of the) students will come to the party
= I believe that some of the students will come to the party.

This appears to be the right result: B's response to A does seem to implicate (12). This effect of scale reversal under negation and other downward-entailing operators, emphasized by several authors (cf., among others, Fauconnier 1975a, 1975b; Atlas & Levinson 1981, Horn 1989) is a welcome result.

2. Embedded Implicatures: a first crack in the Gricean picture

As mentioned, the goal of this article is to challenge the 'neo-Gricean' approach to SIs. We use the term "neo-Gricean" to characterize theories that attempt to derive SIs from Grice's maxims of conversation, generally supplemented with the notion of *scale*, and view SIs as resulting from a reasoning process about speakers' intentions (as in Horn 1972, 1989: Fauconnier 1975a, 1975b, and Levinson 1983, but also, more recently, Spector 2003, 2006, 2007b; Sauerland 2004; van Rooij & Schulz 2004, 2006). Chierchia (2004), in an article which began circulating in 2000, partly building on ideas by Landman (1998), challenged this neo-Gricean approach on the basis of embedded scalar implicatures, and concluded that SIs are derived by means of compositional rules which apply recursively to the constituents of a given sentence (see also Récanati 2003 for a version of this position). Chierchia (2004) thus argues for a *grammatical approach* to scalar implicatures. Several works reacted to this proposal by refining the neo-Gricean approach so as to enable it to account for some of Chierchia's empirical observations. (See for instance Spector 2003, 2006, 2007b; Sauerland 2004; van Rooij & Schulz 2004, 2006; Russell 2006. Horn 2006 is a recent assessment of several aspects of this dispute, from the neo-Gricean standpoint. See also Geurts 2009 for a nuanced defense of the globalist view.)

We are not going to review this literature in details; our goal here is to present what seem to us to be some of the most compelling arguments for a grammatical approach to scalar implicatures. The arguments (some of which are new) will be based on the existence of embedded SIs, and on a variety of additional considerations

We will begin in section 2 with an illustration of what a grammatical approach to SIs might look like, followed by a preliminary argument in favor of such an approach, based on a sub-case of the generalization mentioned in the introduction, namely that embedded implicatures are possible in downward entailing and non-monotonic contexts (non-monotonic operators are operators which are neither upward- nor downward-entailing). Section 3 will provide a detailed new argument for the existence of embedded implicatures in upward entailing contexts. Finally, section 4 will review other arguments that have been recently given for a grammatical approach to SIs.

2.1. Exhaustification as a grammatical device

Does Grice's approach, emended as proposed by Horn, provide us with a framework in which SIs may be properly understood? Horn's move helps us get around the problem

raised by the appeal to relevance; but the epistemic step remains unaccounted for: reasoning via the maxims about the speaker's intentions gets us at best from something like (13a) to (13b):

(13) a. John or Bill will show up
b. The speaker has no evidence that they both will show up

For the time being, let us simply stipulate that the epistemic step can take place and that an utterance of (13a) could be understood as conveying something such as (14a), which we represent as (14b):

(14) a. John or Bill and not both will show up
b. O_{ALT}(John or Bill will show up)

For concreteness, we may assume that if the alternatives are active (and hence the set ALT is non empty), such alternatives are obligatorily factored into meaning via O. Otherwise, if the alternatives are not active, the plain unenriched meaning is used, and no SI comes about. (see also section 4.1.)

So far, we have regarded our silent *only* as way to express, in compact form, Gricean reasoning of the type exemplified in (3). However, a different interpretation of *O* is possible. One might imagine that a silent *only* can be present in the sentence's logical form, and that the scalar implicatures of a given sentence S are in fact *logical entailments* of S when construed as corresponding to an LF in which a silent *only* occurs and takes maximal scope. In this case, exhaustification would be more than just a way of expressing Gricean reasoning compactly. It would become a grammatical device. What lends preliminary plausibility to this interpretation is the observation, stemming from Horn's work, that grammar seems to constrain (via a specification of lexical scales) the set of alternatives.

Before further elaboration, we should state explicitly how our operator O is interpreted. Given a sentence S and a set of alternatives, O_{ALT}(S) expresses the conjunction of S and of the negations of all the members of ALT that are not entailed by S. Equivalently, it states that the only members of ALT that are true are those entailed by S:

(15) $||O_{ALT}(S)||^w = 1$ iff $||S||^w = 1$ and $\forall \phi \in ALT\ (\phi(w) = 1 \rightarrow ||S|| \subseteq \phi)$

In section 3.4.2., we will need to modify this definition, but for the time being (15) is sufficient. Quite often in this article, we'll omit the subscript *ALT* (we make the simplifying assumption that ALT is fully determined by the scalar items that occur in the scope of O). A word of caution is in order: O, so defined, is not exactly equivalent to *only*, for *only* is usually assumed to trigger various presuppositions which *O*, according to the above entry, does not. Note also that the above definition does not encode any direct relationship between *O* and focus marking (see Fox & Katzir 2009).

We can now properly address the main issue of the present article. So far, we have been discussing implicatures that occur in unembedded contexts (like our original case in (3)). And even when we consider cases where implicature triggers (Horn Scale members, a.k.a. scalar items) occur in an embedded position, as in (11) and (12), the relevant implicatures appear to be computed at the root level. This is in keeping with Grice's insight that

implicatures arise by reasoning on speakers intention given a particular speech act (i.e. a whole utterance).

The question we would like to investigate now is whether SIs are *always* computed at the root level. If Grice is right, it should indeed be so. In this respect, the view we are developing that implicatures arise via something like a covert use of *only*, suggests that a different answer to the question might be right. For there is no *a priori* reason for thinking that covert uses of *only* are restricted to the root. If such an operator exists, it is unclear what should prevent its deployment at embedded scope sites. However, whether we are right or wrong on how SIs come about, the issue of whether they can systematically arise in embedded positions clearly deserves close inspection.

Summing up, the Gricean view, emended *à la* Horn with grammatically based constraints on scales, and with the extra assumption of an epistemic step, is able to derive basic SIs. However, such an approach seems to clearly predict that SIs are a root, post-compositional phenomenon. This prediction seems to be confirmed in cases of 'scale reversal' such as those in (11)–(12). The question is whether it withstands further scrutiny.

The rest of this section argues that it does not, based on the existence of so-called 'intrusive' implicatures (see Levinson 2000), i.e., cases where a scalar item retains its 'strengthened' meaning under the scope of a downward entailing (DE) or non-monotonic (NM) operator. In section 3, we'll turn to SIs embedded in UE contexts. As we will see, they are expected to be harder to detect. Nonetheless, we will present various tools that will allow us to see very clear consequences of the presence of such implicatures.

2.2. Implicatures embedded in DE and NM contexts

Sometimes scalar items receive an enriched interpretation under the scope of negation. Examples that seem to force such an enrichment are the following.

(16) a. Joe didn't see Mary or Sue; he saw both.
b. It is not that you *can* write a reply. You must.
c. I don't expect that some students will do well, I expect that all students will.

The first example in (16a) receives a coherent interpretation only if the embedded *or* is interpreted exclusively. Similarly, the modal in (16b) has to be interpreted as *can though need not*, and the quantifier *some* in (16c) as *some though not all*. For all the sentences in (16), in other words, it looks as if the implicature gets embedded under the negative operator. In our notation, the LF of, e.g., (16a) could be represented as:

(17) not O_{ALT}(John saw Mary or Sue)

Examples of this sort have been widely discussed in the literature (especially in Horn 1985, 1989); they seem to require either focal stress on the implicature trigger and/or a strong contextual bias. Horn argues that cases of this sort constitute metalinguistic uses of negation, i.e., ways of expressing an objection not to a propositional content but to some other aspect of a previous speech act. The range of possible speaker's objection can be very broad and concern even the choice of words or the phonology.

(18) You don't want to go to [leisister] square, you want to go to [lester] square

In particular, with sentences like (16), the speaker objects to the choice of words of his interlocutor, presumably for the implicatures they might trigger.

While the phenomenon of metalinguistic negation might well be real (if poorly understood) there are other examples of DE contexts not involving negation that seem to require embedded implicatures, as noted by Horn himself (Horn 1989). In what follows, we will consider several such cases, modelled mostly after Levinson (2000). To begin with, consider the contrast in (19).

(19) a. If you take salad or dessert, you'll be really full.
b. If you take salad or dessert, you pay $ 20; but if you take both there is a surcharge.

The most natural reading of (19a) involves no implicature (*or* is construed inclusively); on the other hand, on the inclusive reading, (19b) would be contradictory. If an indicative conditional sentence is analyzed as a material or a strict conditional, a coherent interpretation, which is clearly possible in the case of (19b), requires an embedded implicature. Let us go through the reasons why this is so. Suppose that in the context where (19b) is uttered, the alternative with *and* is active. Then, there may be in principle two sites at which the implicature is computed. Using our notation, they can be represented as follows:

(20) a. O_{ALT}(if you take salad or dessert, you pay $ 20)
b. if O_{ALT}(you take salad or dessert), you pay $ 20

If the option in (20a) is taken, the relevant alternative set would be as in (21b):

(21) a. If you take salad or dessert, you pay $ 20
b. ALT = {If you take salad or dessert, you pay $ 20, If you take salad and dessert, you pay $ 20}
c. O_{ALT} (if you take salad or dessert, you pay $ 20) = 1 iff
If you take sale or dessert, you pay $ 20 $\wedge \forall p \in$ ALT such that p is not entailed by 'if you take salad or dessert, you pay $ 20', p is false.

Since the assertion (21a) entails all its alternatives, the truth conditions of (20a) wind up being the same as those of (21a) (i.e., no implicature comes about – cf. the computation in (21c)). And as already noted, this reading is too strong to be compatible with the continuation in (19b). So the LF in (20a) cannot account for the coherent interpretation of (19b). On the other hand, if the implicature is computed at the level of the antecedent, as in (20b), we get the equivalent of:

(22) If you take salad or dessert and not both, you pay $ 20

The truth conditions of (22) are weaker than those of (20a), and this makes them compatible with the continuation in (19b). Thus, the fact that sentences such as (19b) are

acceptable seems to constitute prima facie evidence in favor of the possibility of embedding SIs.

One might wonder whether an alternative, non-monotonic analysis of conditionals, e.g., that of Stalnaker (1968) or Lewis (1973) could provide an account which would not rely on embedded implicatures. Although space limitations do not allow us to discuss this in any detail, we would like to point out that any such account would also have to explain the different behavior of conditional sentences in which the antecedent does not license a scalar implicature in isolation. (#*If you take salad or dessert or both, you pay exactly $ 20; but if you take both there is a surcharge.*).

This phenomenon seems to be quite general. Here are a few more examples involving the antecedents of conditionals, as well as further DE contexts like the left argument of the determiner *every*.

(23) a. If most of the students do well, I am happy; if all of them do well, I am even happier
b. If you can fire Joe, it is your call; but if you must, then there is no choice
c. Every professor who fails most of the students will receive no raise; every professor who fails all of the students will be fired. (M. Romero p.c.)

It should be noted that these examples can be embedded even further.

(24) John is firmly convinced that if most of his students do well, he is going to be happy and that if all of them will do well, he'll be even happier.

(25) Every candidate thought that presenting together his unpublished papers and his students evaluation was preferable to presenting the one or the other.

Without adding implicatures at a deeply embedded level, all of these examples would be contradictory. For instance, in (24) the implicature is embedded within the antecedent of a conditional, which is in turn embedded under an attitude verb. In (25), a coherent interpretation is possible only if *or* (in *the one or the other*) is interpreted as exclusive; note indeed that structures of the form *A is preferable to B* are perceived as contradictory when A entails B, as evidenced by the oddness of, e.g., *Having a cat is preferable to having a pet*. Thus if disjunction were inclusive in (25), it should sound odd as well.

A similar argument can be replicated for non-monotonic contexts. Consider for instance:

(26) Exactly two students wrote a paper or ran an experiment

It seems relatively easy to interpret *or* in (26) exclusively. As a first step towards seeing that this is the case, it is useful to note that the truth conditions associated with the exclusive and inclusive construals of *or* under the scope of a non-monotonic quantifier as in (26) are logically independent. For example, in a situation in which one student writes a paper and another writes a paper and also runs an experiment (and nobody else does either), the sentence is true on the inclusive construal of *or*, but false on the exclusive construal. On the other hand, in a situation in which one student only writes a paper, another only runs an experiment and other students do both, the inclusive interpretation of *or* in

(26) is falsified; in such a scenario, (26) is only true on the (embedded) exclusive construal. Now the exclusive reading can be easily observed when we consider environments in which it forced:

(27) Exactly two students wrote a paper or ran an experiment. The others either did both or made a class presentation.

For (27) to be coherent, the implicature must be computed under the scope of *exactly two*. Cases of this sort are pretty general and can be reproduced for all scalar items: sentence (28) below must be interpreted as 'exactly three students did most though not all of the exercises'

(28) Exactly three students did most of the exercises; the rest did them all.

Taking stock, we have discussed a number of example sentences involving a variety of DE and NM contexts. Such sentences appear to have coherent interpretations that can only be derived if an implicature is computed at an embedded level (i.e., within the scope of a higher verb or operator). It should be noted that focal stress on the scalar item often helps the relevant interpretation. From our point of view, this is not surprising. The mechanism we have sketched for implicature calculation is, in essence, covert exhaustification, one of the phenomena triggered by focus. More generally, for the time being, we make no claim as to the frequency or marked status of embedded implicatures (but see our discussion in section 4.6 below). Our point is simply that they can and do naturally occur and that there are ways in which embedded implicatures can be systematically induced. This fact seems to be incompatible with the claim that SIs are a postcompositional semantic process, as the Gricean or Neo-Gricean view would have it. Of course, to establish our claim fully, we would like to be able to show that SIs can also be embedded in UE contexts, which, as we will see shortly, calls for more sophisticated methods.

3. A new argument for embedded implicatures in UE contexts: Hurford's constraint

If embedded implicatures exist, we expect many sentences to be multiply ambiguous, depending on whether an SI is computed in a given embedded position or not. Furthermore, in UE contexts, the various readings that are predicted are all stronger than the 'literal' reading, and in many cases, the presence of an embedded implicature yields a reading which is stronger than both the literal reading and the reading that results from applying the standard Gricean reasoning to the sentence as a whole (in our terms, the reading that results from applying *O* to the whole sentence). So, it may prove hard to establish the possibility of embedded implicatures by mere inspection of truth-conditions (since if a certain reading R1 entails another reading R2, there can be no situation where R1 is accepted as true and R2 is not; see, e.g., Meier & Sauerland 2008 for a related methodological discussion).

In order to circumvent this difficulty, it would be useful to have constructions in which only one of the potentially available readings is licensed. In this section, we are going to

argue that such constructions exist: we'll show that some sentences will *have* to contain a local exhaustivity operator in a UE context for a certain constraint (Hurford's constraint) to be met. The general line of argumentation is presented in a more detailed way in Chierchia, Fox & Spector (2009).

3.1. Hurford's constraint

Hurford (1974) points to the following generalization:

(29) Hurford's constraint (HC): A sentence that contains a disjunctive phrase of the form *S or S'* is infelicitous if *S* entails *S'* or *S'* entails *S*.

This constraint is illustrated by the infelicity of the following sentences:

(30) a. # Mary saw an animal or a dog.
b. # Every girl who saw an animal or a dog talked to Jack.

However (31) below, which is felicitous, seems to be a counterexample to HC:

(31) Mary solved the first or the second problem or both

If *or* is interpreted inclusively, then clearly (31) violates HC, since 'Mary solved both problems' entails 'Mary solved the first problem or the second problem'. On the basis of such examples, Hurford reasoned that *or* has to be ambiguous, and that one of its readings is the exclusive reading. On an exclusive construal of the first disjunction in (31), the sentence no longer violates HC. Gazdar (1979) noticed other cases where HC appears to be obviated, such as (32):

(32) Mary read some or all of the books

By analogy with Hurford's reasoning about disjunction, one might conclude that *some* is ambiguous as well, and means *some but not all* on one of its readings. But Gazdar argued that multiplying lexical ambiguities in order to maintain HC misses an obvious generalization. Gazdar, thus, proposed to weaken Hurford's generalization in the following way:

> Gazdar's generalization: A sentence containing a disjunctive phrase *S or S'* is infelicitous if *S* entails *S'* or if *S'* entails *S*, unless *S'* contradicts the conjunction of *S* and the implicatures of *S*. (Note that Gazdar's generalization is asymmetric, i.e., allows for cases where *S'* entails *S* but not for cases where *S* entails *S'*. We do not address this point here. See also Singh 2008)

In both (31) and (32), the SI associated with the first disjunct (*Mary did not solve both problems*, *Mary did not read all of the books*) contradicts the second disjunct, and so both sentences are predicted to be felicitous by Gazdar's generalization. Gazdar himself did not offer an account of his generalization. We will provide such an account, along the following lines:

- HC is correct as originally stated
- All apparent violations of HC involve the presence of an implicature-computing operator within the first disjunct, ensuring that HC is met – hence the presence of a 'local implicatures'

In a sense, we extend Hurford's original account based on ambiguity to all scalar items, even though we do not assume a *lexical* ambiguity, and instead derive the various readings with our implicature-computing operator. It should be clear that we derive something close to Gazdar's generalization: suppose S2 entails S1; then 'S1 or S2' violates HC; yet 'O_{ALT}(S1) or S2' may happen to satisfy HC; this will be so if S1 together with its implicatures is no longer entailed by S2, which will be the case, in particular, if S2 contradicts S1 together with its implicatures

However, as we will show in the next subsections, our proposal turns out to make very precise and new predictions in a number of cases – predictions that do not fall out from anything that Gazdar has said.

3.2. Forcing embedded implicatures

Gazdar's generalization, as such, does not make any particular prediction regarding the *reading* that obtains when there is an apparent violation of HC. But consider now the following sentence (in a context where it has been asked which of a given set of problems Peter solved):

(33) Peter either solved both the first and the second problem or all of the problems.

In the absence of an exhaustivity operator, (33) would violate HC, since solving all of the problems entails solving the first one and the second one. And (33) would then be equivalent to (34):

(34) Peter solved the first problem and the second problem.

Therefore, we predict that an exhaustivity operator has to be present, with the effect that (33)'s logical form is the following:

(35) O_{ALT}(Peter solved the first problem and the second problem) or he solved all of the problems

Recall that the meaning of our operator is supposed to be – as a first approximation – the same as that of *only*. If we are right, the meaning of (33) could be given by the following paraphrases (which are themselves equivalent):

(36) a. Peter only solved the first problem and the second problem, or he solved all of the problems
b. Either Peter solved the first problem and the second problem and no other problem, or he solved all the problems

It turns out that this is indeed the only possible reading of (33). In other words, (33) is clearly judged *false* in a situation in which Peter solved, say, the first three problems and

no other problem (out of a larger set). So (33) seems to be a clear case of an embedded implicature. What distinguishes (33) from cases like (31) and (32) is that exhaustifying the first disjunct has a semantic effect in this case.

In Chierchia, Fox & Spector (2009), we discuss more complex cases in which exactly the same logic is at play. But the general form of the argument should be clear: if HC is correct as originally formulated, then, in some cases, the only way to satisfy HC is to insert O locally at an embedded level, and this gives rise to readings which turn out to be the only possible ones. The next section develops a similar, though more complex, argument.

3.3. HC and recursive exhaustification

In this section, we are going to show that even in cases where the obligatory presence of embedded O does not have any direct effect on the literal truth-conditions of a sentence, it nevertheless has consequences that can be detected by investigating the *implicatures* of the sentence in question (or lack thereof). In our terms, the presence of the embedded implicature-computing operator turns out to have a truth-conditional effect when the sentence is itself embedded under another implicature-computing operator. First, we'll look at the interpretation of disjunctions of the form 'A or B or both' in the scope of necessity modals (3.3.1). Then we'll offer an account of the 'cancellation effect' triggered by *or both* in non-embedded contexts (3.3.2).

3.3.1. *Or both* in the scope of necessity modals

Consider the following two sentences:

(37) We are required to either read *Ulysses* or *Madame Bovary*

(38) We are required to either read *Ulysses* or *Madame Bovary* or both

Both these sentences implicate that we are not required to read both *Ulysses* and *Madame Bovary*. At first sight, they do not seem to trigger different implicatures. But upon further reflection, they, in fact, do. Suppose that we are actually required to read *Ulysses* or *Madame Bovary* and that we are *not* allowed to read both of them. Then (37) would be an appropriate statement, while (38) would not. Sentence (37), on its most natural reading, is silent as to whether or not we are allowed to read both novels. But (38) strongly suggests that we are allowed to read both novels. So (38) seems to trigger an implicature that (37) does not.

This is further confirmed by looking at the following dialogues:

(39) A: We are required to either read *Ulysses* or *Madame Bovary*
B: No! we have to read both

(40) A: We are required to either read *Ulysses* or *Madame Bovary*
B: ## No! We are not allowed to read both

(41) A: We are required to either read *Ulysses* or *Madame Bovary* or both
B: No! We are not allowed to read both

(39) serves to set the stage. It shows what we've already seen in section 2, namely that an implicature can be embedded below negation: Speaker B, in (39), is objecting, not to the literal meaning of A's utterance but to an implicatures of this utterance, namely, that we are not required to read both novels. In section 2 we've argued that phenomena of this sort argue for embedded implicatures, but this is not important in the current context. What is important here is that the relevant phenomena could be used to investigate the implicatures of various utterances, and as such they distinguish between (37) and (38). In (40), B's reply is clearly deviant. This shows that A's utterance in (40) (= (37)) does *not* implicate that we are allowed to read both novels; indeed, if A's sentence did trigger this implicature, then B's objection would be perfectly felicitous, since it would count as an objection to an implicature of A's utterance. But now notice the contrast with (41). B's objection in (41) is completely natural, and hence confirms our claim that in (41), A's sentence (= (38)) does implicate that we are allowed to read both novels.

How are these facts to be explained? Why does (38) have an implicature that (37) doesn't? Note that (modulo the effects of matrix *O*) (37) and (38) have the same truth-conditions. Yet they trigger different implicatures. We are going to show that this phenomenon is in fact entirely expected from our perspective. The implicatures associated with (37) and (38) are, in fact, instances of the following generalization:

(42) A sentence of the form $\Box$*(A or B)* triggers the following implicatures (where $\Box$ stand for any modal operator with universal force): $\neg\Box A$, $\neg\Box B$

To illustrate this generalization let us begin with (37), repeated here as (43), which implicates that we have a choice as to how to satisfy our obligations.

(43) We are required to either read *Ulysses* or *Madame Bovary*

The reading of (43), 'pragmatically strengthened' on the basis of the generalization in (42), is given in (44a), which is equivalent to (44b) (because 'being not required to do X' is equivalent to 'being allowed not to do X') and then to (44c).

(44) a. We are required to either read *Ulysses* or *Madame Bovary* and we are not required to read *Ulysses* and we are not required to read *Madame Bovary*.
b. We are required to either read *Ulysses* or *Madame Bovary*, we are allowed not to read *Ulysses* and we are allowed not to read *Madame Bovary.*
c. We are required to either read *Ulysses* or *Madame Bovary*, and we are allowed to read either one without reading the other

Now, let's see the consequences of this generalization for the case of (38) ("We are required to read *Ulysses* or *Madame Bovary* or both"), schematized as follows:

(45) $\Box$ [O_{ALT}(A or B) or (A and B)]

(45) is predicted to implicate the following:

(46) a. $\neg\Box(O_{ALT}(A \text{ or } B))$
b. $\neg\Box$(A and B)

The end-result is the following proposition, which indeed corresponds to the most natural reading of (38):

(47) We are required to read *Ulysses* or *Madame Bovary*, we are not required to read only one of the two novels, we are not required to read both novels.

From (47) it follows that we are allowed to read both novels, which was the desired result. So far we have shown that given the generalization in (42), the observed interpretation of (38) follows directly from the assumption that O is present in the first disjunct. Of course, it is also desirable to understand why generalization (42) should hold. It turns out that the exhaustivity operator as we have defined it so far (adding the negations of all non-weaker alternatives) can derive all the inferences that we observed, provided we now assume that the scalar alternatives of a disjunctive phrase *X or Y* include not only the phrase *X and Y* but also each disjunct *X* and *Y* independently. If so, then (43) has the following alternatives (using now the sign '□' to abbreviate 'we are required to', 'U' to abbreviate 'read *Ulysses*', and 'MB' to stand for 'read *Madame Bovary*'):

(48) ALT((43)) = {□(U or MB), □U, □MB, □(U and MB)}

To get the 'strengthened meaning' of (43), we simply assume that (43) gets exhaustified yielding the Logical Form in (49), which involves adding the negation of each of its alternatives that is not weaker than the assertion. The result is:

(49) $O_{ALT}((43)) = O_{ALT}\Box(U \text{ or } MB)) = \Box(U \text{ or } MB) \wedge \neg\Box U \wedge \neg\Box MB$

Let us now go back to our original example (38) to see how this assumption plays out:

(50) a. We are required to either read *Ulysses* or *Madame Bovary,* or both
b. O_{ALT} (□(O_{ALT}(U or MB) or (U and MB)))

The logical form of (50a) must be (50b), where the embedded exhaustivity operator is forced by HC and the matrix one exhaustifies the entire sentence. Given our new assumptions about the alternatives of a disjunctive phrase, (50)'s alternatives include, among others, '□(O_{ALT}(U or MB)' (i.e., "we are required to read either *Ulysses* or *Madame Bovary* and we are forbidden to read both") and '□(U and MB)'. Focussing (for simplicity) on just these two alternatives, (50b) tantamounts to

(51) a. $\Box(U \text{ or } MB)) \wedge \neg\Box(O_{ALT}(U \text{ or } MB)) \wedge \neg\Box(U \text{ and } MB)$
b. We are required to either read *Ulysses* or *Madame Bovary,* and we are not required to read only one of them, and we are not required to read both
c. We are required to read *Ulysses* or *Madame Bovary* and we are allowed to read both of them and we are allowed to read only one of them

This is exactly what we wanted to derive. Let us sum up what has been shown. The obligatory presence of an exhaustivity operator applying to the first disjunct in sentences like (38), together with the assumption that each member of a disjunctive phrase contributes to the alternatives of the disjunctive phrase, immediately predicts that (38), though

equivalent to (37), has more alternatives. One of these alternatives, crucially, contains an exhaustivity operator. Due to this additional alternative, (38) triggers an inference that (37) does not (the inference that we are not required to read *only* one of the two novels), a prediction that appears to be correct.

3.3.2. Symmetric alternatives and Innocent Exclusion

In the next sub-section we will show that our perspective enables us to derive the fact that disjunctions of the form *A or B or both*, or *some or all*, do not trigger any SI if they are not embedded (a fact that is sometimes described as a 'cancellation' effect). But first we need to elaborate a bit on the idea that the disjuncts are alternatives of a disjunctive sentence. This idea, which seems to be needed for a variety of purposes in addition to the one mentioned above (see Sauerland 2004; Spector 2003, 2007b and Fox 2007), doesn't follow from the Horn set for disjunction given in (9d). One way of modifying the Horn set for binary connectives along the lines we have discussed is suggested by Sauerland (2004), namely expanding such a set to include the abstract elements L and R, where *p L q* is equivalent to *p* and *p R q* is equivalent to *q*:

(52) **Sauerland's alternatives for disjunction:** {*or*, *L*, *R*, *and*}

(For alternatives see Spector 2006, Katzir 2007, and Alonso-Ovalle 2005). The obvious problem with the proposal that the disjuncts are among the alternatives of a disjunctive statement is that our operator *O*, as we have defined it in section 2, now yields a contradiction when it applies to a simple disjunctive sentence: according to our definition, *O(p or q)* states that the disjunctive sentence is true and each of its alternatives (none of which it entails) is false. In particular, it would now follow that each of the disjuncts is false, but this of course contradicts *(p or q)*. The problem arises whenever the set of alternatives has a property that we might call 'symmetry', i.e., whenever it contains two or more alternatives that can be excluded separately but not together (e.g., *p or q* is consistent with the exclusion of *p* or with the exclusion of *q*, but the moment both are excluded the result is contradictory).

In order to correct for this problem, we have to revisit the question of which alternatives get to be excluded. Our previous definition stated that entailed alternatives don't get excluded. But, why? The reason seems rather obvious. Exclusion of weaker alternatives would lead to an automatic contradiction. But, as we've just seen, there are other exclusions that would lead to a contradiction, namely the simultaneous exclusion of two symmetric alternatives. We might therefore suggest that *O* be modified to avoid contradictions in situations of symmetry. In particular, given a sentence S and a set of alternatives C, we could define a set of innocently excludable alternatives, I-E(S,C), as the set of sentences that can all be false while S is true. We take this definition from Fox (2007), who took his inspiration from Sauerland's algorithm for the computation of SIs. (See also Spector 2003, 2007b and van Rooij & Schulz 2004, 2006 for related proposals.)

The definition of I-E(S,C) proceeds in two steps. First we look at the biggest subsets C' of C such that $\{\neg S' : S' \in C\} \cup \{S\}$ is a consistent set of propositions. Each such set corresponds to a way of excluding as many alternatives as possible in a consistent manner (i.e., consistent with the basic assertion). For instance, if S is of the form *p or q* and C is {*p or q, p, q, p and q*}, there are two sets which meet this condition, namely C' = {*p, p and q*} and

C'' = {*q, p and q*}. This is so because (a) {*p or q, ¬p, ¬(p and q)*} is consistent (every member of this set is true if q is true and p is false), and so is {*p or q, ¬q, ¬(p and q)*}, and (b) C' and C'' are, furthermore, maximal such sets, since the only bigger subset of C that does not include S is {*p, q, (p and q)*}, whose members cannot be consistently denied together if S itself (*p or q*) is asserted. The set of innocently excludable alternatives, I-E(S,C), is then constructed by collecting the alternatives which belong to *every* such maximal set. These are the alternatives which we know can be excluded safely (i.e., consistently), irrespective of which other alternatives have been already excluded. O is then defined in terms of innocent exclusion: when applied to a sentence S relative to a set of alternatives C, it returns the conjunction of S and the negations of all the innocently excludable members of C (i.e., of all the members of I-E(S, C)). This definition of O is summarized below:

(53) a. $O_C(S)$ is true iff *S* is true and for any S' in I-E(S,C): S' is false
b. I-E(S,C) is the intersection of maximal excludable alternatives of C given S.
c. M⊆C is an excludable alternative of C given S, if the conjunction of S and the negation of all members of M is consistent.
d. M is a maximal excludable alternative of C given S if M is an excludable alternative of C given S and there is no superset of M included in C which is an excludable alternative of C given S.

The result of applying *O*, at times, could be rather difficult to compute, but the notion itself is rather simple. All stronger alternatives are excluded unless there is symmetry, in which case all alternatives that do not partake in symmetry are excluded. However, in most cases we do not need consider the definition in (53). Specifically, whenever using our initial definition of O does not result in a contradiction, we can continue to use it, and the reason is simple. If using our previous definition yields a non-contradictory result, then the set of innocently excludable alternatives is precisely the set of sentences in C not entailed by S. We will, thus, continue to use our initial definition of *O* whenever the result is consistent.

3.3.3. Cancellation

Consider the two following sentences:

(54) a. John bought some of the furniture.
b. John bought some or all of the furniture.

Since the early days of theorizing on the nature of SIs, the fact that (54b) is not associated with the same SIs as the simple sentence in (54a) has been highly problematic. We will see that the problem is eliminated the moment the role of embedded exhaustification is understood.

The two sentences in (54) are equivalent. Under the Neo-Gricean Theory, as well as the grammatical alternative we are considering, the only way two equivalent sentences can be systematically associated with different SIs is if they have different scalar alternatives. The problem is that, without an embedded exhaustivity operator, (54a) and (54b) have equivalent scalar alternatives. Consider indeed the scalar items in (54b): *some*, *all*, and *or* with the following Horn Sets.

(55) a. Connectives: {*or*, *L*, *R*, *and*}
b. Quantifiers: {*some*, *all*}

We thus get the following sentential alternatives:

(56) a. Alternatives for (54a):
{John bought some of the furniture, John bought all of the furniture}
b. Alternatives for (54b):
{John bought some of the furniture, John bought some or all of the furniture, John bought some or some of the furniture, John bought some and some of the furniture,

John bought all of the furniture, John bought all or all of the furniture, John bought some and all of the furniture, John bought all and all of the furniture}

Although there are more alternatives in (56b) than in (56a), they divide into two sets of equivalent alternatives (separated by a line). One set is equivalent to the *some* alternative in (56a) and the other to the *all* alternative. It follows that the same SI is predicted. The same holds if we use a more traditional Horn set, hence this recalcitrant problem in the Neo Gricean tradition is fairly independent of the particular perspective on alternatives we've been taking.

Interestingly the problem is obviated the moment *O* applies to the first disjunct. We now have another alternative, namely the first disjunct *O(John bought some of the furniture)*, which states that John bought some but not all of the furniture. This alternative is symmetric relative to the alternative based on *all*. One cannot exclude both, and, hence neither is innocently excludable. There are thus no innocently excludable alternatives, and matrix application of *O* is vacuous, precisely the result we want. We now see clearly why the addition of *or both* in sentences such as (54) has the effect of canceling the exclusive implicature typically associated with *or*.

3.4. Conclusion

In this section, we have offered a theory of the interaction of SIs and Hurford's Constraint. HC provides us with a way of probing the existence of embedded implicatures. Sometimes O has to be embedded for HC to be satisfied. Moreover, once we allow for embeddings of O, as we sometimes have to, we make a host of complex predictions that we have pain stakingly laid out and argued to be borne out. This further corroborates our main hypothesis, namely that scalar implicatures can be freely embedded.

4. Further cracks in the Gricean picture

As we mentioned at the very beginning, it is not clear how the availability of embedded scalar implicatures could be made consistent with a Gricean approach to SIs. More specifically, the facts suggest that SIs are not pragmatic in nature but arise, instead, as a consequence of semantic and syntactic mechanisms, which we have characterized in terms of the operator, O.

In this concluding section we would like to mention a few other observations that support this conclusion. Our discussion will be extremely sketchy given the limited space we have available. Nevertheless, we will try to introduce the pertinent issues and will refer the reader to relevant literature.

4.1. Obligatory SIs

A property that is commonly attributed to SIs is their optionality, sometimes referred to as "cancelability":

(57) a. John did some of the homework. In fact, he did all of it.
b. John did the reading or the homework. He might have done both.

The first sentences in (57a) and (57b) would normally lead to SIs (*not all* and *not both*, respectively). But these SI are not obligatory, else the continuation would lead to a contradiction.

The optionality observed in (57) is a necessary attribute of SIs from the (neo-)Gricean perspective. According to the latter, SIs are not automatic but rather follow from two assumptions that don't always hold, namely the assumption that the speaker is opinionated about stronger alternatives (which justifies the epistemic step alluded to in section 1.2.), and the assumption that the stronger alternatives are contextually relevant. The fact that these assumptions are not necessarily made in every context explains optionality.

This optionality is also captured by our grammatical mechanism. Given what we've said up to now, there is nothing that forces the presence of the operator *O* in a sentence containing a scalar item. Optionality is thus predicted, and one can capture the correlation with various contextual considerations, under the standard assumption (discussed in the very beginning of this article) that such considerations enter into the choice between competing representations (those that contain the operator and those that do not). An alternative way of capturing optionality is to assume that there is an optional process that activates the alternatives of a scalar item, but that once alternatives are active, SIs are obligatory (see Chierchia 2006 for an implementation). Under the latter view, what is optional is the activation of alternatives; if alternatives are activated, they must be factored into meaning via O.

This second option has a consequence. If under certain circumstances scalar alternatives have to be actived, *obligatory* SIs are expected to arise. The claim that this is so is, in fact, implicitly present in Krifka (1995) and Spector (2007a), and has been explicity defended in Chierchia (2004, 2006) and Magri (2009). We cannot go over all of the arguments and will narrow our attention to an argument made by Spector (2007a) in the domain of plural morphology.

Consider the contradictory status of the utterance in (58).

(58) #John read (some) books; maybe he read only one book.

This contradiction suggests that the first sentence *John read (some) books* (on both its variants) is equivalent to the statement that there is more than one book that John read. However, assuming this as the basic meaning doesn't account for its behavior in downward entailing environments. Consider the sentences in (59). Their interpretation (at least the one that immediately springs to mind) is stronger than what would be predicted under the putative meaning for the first sentence in (58).

(59) a. John didn't read books.
b. I don't think that John read (some) books.

To see this, focus on (59a). The sentence (under its most natural interpretation) is false if John read exactly one book. The same point can be made for (59b), and both points are illustrated by comparison with sentences in which *John read (some) books* is substituted by a sentence that clearly has the putative meaning, namely *John read more than one book*:

(60) a. John didn't read more than one book.
b. I don't think that John read more than one book.

We seem to be facing something close to a paradox. In order to account for the meaning of (58), *John read (some) books* must have a strong meaning, namely that John read *more* than one book, and in order to account for the meaning of (59), it must have a weaker meaning, namely that John read *at least* one book.

Building on suggestions made by Sauerland (2003), Spector argues that the basic meaning is the *at least one* meaning, and that the stronger meaning (i.e., the *more than one* meaning) is a scalar implicature. Explaining how the strong meaning is derived is a rather complicated matter, which we will have to skip in this context. What is important, however, is that implicatures can easily disappear in downward entailing contexts, which accounts for the most natural readings of (59a) and (59b). As explained in section 2.1., the fact that scalar items need not retain their strengthened meaning in DE contexts is an automatic consequence of the neo-Gricean approach. It is also an automatic consequence of the grammatical perspective that we are advocating, since an exhaustivity operator (even if obligatorily present) need not be inserted below a DE operator. Let us see how things work in the case at hand.

If we assume that the plural morpheme *pl* makes it *obligatory* to insert the operator *O* in some syntactic position that c-commands *pl*, we expect the following: in a simple, non-embedded context, *O* can only be inserted just above the plural morpheme, which gives rise to an *at-least two* reading (as demonstrated in Spector's work); but if *pl* occurs in the scope of a DE-operator, more options are available; in particular, *O* may be inserted at the top-most level, i.e., above the DE-operator, in which case no implicature will arise (because inserting *O* at the top-most level always gives rise to the reading that is predicted by the pragmatic, neo-Gricean, approach). In the case of the plural morpheme, we therefore predict that the *at least two*-reading *can* disappear in DE-contexts, while it is obligatory in non-embedded UE contexts. This will generally be a property of *obligatory scalar implicatures*: the strengthened meaning of an item that *must* occur under the scope of O will be the only option in non-embedded UE contexts, but will appear to be optional in DE contexts. In the case of the plural morpheme, this seems to be just right. Notice that, as expected, the *at-least-two* reading actually *can* be maintained in DE-contexts, with the appropriate intonation pattern:

(61) Jack may have read *one* book; but I don't think he has read book*s*.

If all this is correct, it means that the implicature generated by plural morphology is obligatory (which is why (58) is contradictory in every context). As mentioned, this cannot be captured under neo-Gricean assumptions but can be made to follow from a grammatical theory that incorporates the operator *O*. Specifically, under the grammatical theory, one can maintain that for plural morphology, in contrast with standard scalar items,

alternatives are automatically activated. Once alternatives are active, they must be associated with the operator *O*. This operator yields the *more than one* interpretation for (58). However, once (58) is embedded under a downward entailing operator (e.g., ((59)), the stronger alternatives are now weaker, and the relevant implicatures can be eliminated.

While our discussion has been extremely sketchy (as promised), we hope that the nature of the argument is clear. Gricean implicatures must be optional/cancelable. But if implicatures are derived by a grammatical mechanism, they are optional only when the mechanism is optional, and that, in turn, may depend on various grammatical factors. A similar argument has been made in other domains. Most famously, Krifka (1995) has argued that negative polarity items are obligatorily associated with alternatives, and that these alternatives yield obligatory implicatures which account for their distributional properties. This argument has been developed by Chierchia (2004) to account for intervention effects (discussed in section 4.3 below) and has been extended to other polarity items in Chierchia (2006).

4.2. Encapsulation

Consider the oddness of the following:

(62) John has an even number of children. More specifically, he has 3 (children).

The source of the oddness is intuitively clear: the second sentence seems to contradict the first sentence. However, it is not trivial to account for the contradiction. The second sentence *John has 3 children* (henceforth just *3*) has an interpretation which is consistent with the first sentence, e.g., an interpretation consistent with John having exactly 4 children, which is, of course, an even number. So, why should the two sentences feel contradictory? If in the context of (62), *3* was required to convey the information that John has exactly 3 children, the contradiction would be accounted for. But what could force this "exactly" interpretation on the sentence?

It is tempting to suggest that the theory of SIs should play a role in the account. If in (62) the implicature is obligatory, then the second sentence would contradict the first. And indeed, as we see in (63), and as we've already seen in the previous sub-section, there are some cases where implicatures are obligatory:

(63) Speaker A: Do you know how many children John has?
Speaker B: Yes, he has 4 children. #In fact, he has 5.

However, it turns out that the Gricean reasoning that has been proposed to account for SIs does not derive the attested contradiction. This is a point that was made in a different context by Heim (1991), and was discussed with reference to (62) and similar examples in Fox (2004) and most extensively in Magri (2009). To understand the argument, it is useful to try to derive the SI, along the lines outlined in section 1, and to see where things break down.

So, let's try. Upon hearing the utterance of *3*, the addressee (h, for hearer) considers the alternative sentences in (64), and wonders why the speaker, s, did not use them to come up with alternative utterances.

(64) a. More specifically, he has 4 children.
b. More specifically, he has 5 children.
c. ...

Since all these (clearly relevant) alternatives are stronger than s's actual utterance, h concludes based on (the assumption that s obeys) the Maxim of Quantity that s does not believe any of these alternatives. i.e., s derives the conclusions in (65), which together with the basic utterance, *3*, can be summarized as (66).

(65) a. $\neg B_s$(John has 4 children).
b. $\neg B_s$(John has 5 children).
c. ...

(66) O_{ALT} [B_s(John has 3 children)].

Now, based on the assumption that s is opinionated with respect to the alternatives in (64), h might take 'the epistemic step' (tantamount to 'neg-raising'), which leads to the conclusions in (67), summarized in (68).

(67) a. $B_s\neg$(John has 4 children).
b. $B_s\neg$(John has 5 children).
c. ...

(68) B_s[O_{ALT} (John has 3 children)].

This conclusion clearly contradicts the first sentence in (62), thus, accounting for the observed phenomenon. We thus seem to have a purely neo-Gricean account of the deviance of (62). But this impression is mistaken, as the following illustrates.

The problem is that we were too quick to derive the conclusions in (65) based on the Maxim of Quantity. It is true that all of the utterances in (64) are *logically* stronger than *3*, but are they all also *more informative*, given the special properties of the immediate context? To answer this question we have to understand what is taken to be true at the point at which *3* is uttered (i.e., after the first sentence in ((62)). If the first sentence in (62) is already taken to be true, i.e., if it is assumed that John has an even number of children, the proposition that John has at least 3 children (the relevant meaning of *3*), and the proposition that John has at least 4 children (the relevant meaning of (64a)) provide *exactly the same information*, namely that John has an even number of children greater or equal to three, i.e., that he has 4 or more children.

So, the Maxim of Quantity does not require s to prefer (64a) to *3*. Therefore, the inference in (65a) does *not* follow from the assumption that s obeys the maxim. Moreover, since (65a) and the second sentence in (62), which uses the number *3*, convey exactly the same information, they are predicted to yield exactly the same SI, which together with the basic contextual meanings amounts to the proposition that John has an even number of children greater or equal to 3, but does not have an even number of children greater or equal to 5, which is, of course, tantamount to saying that John has exactly 4 children. So the only implicature we get by employing this purely Gricean reasoning fails to make (62) incoherent.

In other words, on closer scrutiny, it turns out that we fail to account for the contradictory nature of (62). The Gricean reasoning predicts that (62) will be just as appropriate as the following:

(69) John has an even number of children. More specifically, he has 4 children.

This is in sharp contrast with what happens if SIs are derived within the grammar, using the operator O. Under such a view, the contradiction is derived straightforwardly. The sentence *3* activates alternatives which are operated on by O 'blindly', as it were, and when this happens we obtain the proposition that John has exactly 3 children, and this proposition directly contradicts the earlier sentence which asserts that John has an even number of children.

To couch it differently, what (62) seems to teach us is that the notion of informativity relevant to SI computation is logical entailment, rather than entailment given contextual knowledge. This means that the module that computes SIs has to be encapsulated from contextual knowledge, which makes sense if the module is (part of) grammar but not if it is (part of) a "central system" used for general reasoning about the world, as Grice envisioned. For further arguments to this effect, see Fox & Hackl (2006) and in particular Magri (2009).

As various examples in Magri (2009) illustrate, our argument does not specifically rely on the use of numerals, which we selected here to simplify the exposition. The general point can be made with other scalar items, even though one has to construct more complicated discourses, such as the following:

(70) Every student, including Jack, solved either none of the problems or all of the problems. #Jack solved some of the problems.

In this case, the second sentence, under its non-strengthened, logical meaning, is compatible with the first (and contextually entails that Jack solved all of the problems). Yet it is felt as contradictory. So even if numerals only had an 'exact' meaning, as suggested by Breheny (2008), our general point would remain.

4.3. Negative polarity items and intervention effects

Negative Polarity Items (NPIs) (e.g., *any*) are generally licensed in downward entailing contexts (Ladusaw 1979, Fauconnier 1975a). However, as pointed out by Linebarger (1987), certain logical operators appear to disrupt this licensing:

(71) a. John didn't introduce Mary_1 to anyone she_1 knows
b. *John didn't introduce $[\text{every woman}]_1$ to anyone she_1 knows.

This intervention effect has been studied extensively in the literature. Among the important observations that have come out of this study is a typology of the logical operators that yield an intervention effects, henceforth intervening operators (cf., among others, Linebarger 1987; Guerzoni 2004, 2006). Compare (71b) to (72), and (73a) to (73b).

(72) John didn't introduce $[\text{a single woman}]_1$ to anyone she_1 knows

(73) a. John didn't talk either to Mary or to any other girl.
b. *John didn't talk both to Mary and to any other girl.

This comparison leads to the conclusion that existential quantification and disjunction do not yield intervention affects, but universal quantification and conjunction do (Guerzoni 2004, 2006). Why should this be the case? Chierchia (2004) suggests that the answer follows from the theory of SIs. We cannot go over the details of the proposal but we can introduce the basic idea. Assume first that licensing of NPIs requires them to be in a DE context. Assume, furthermore, that SIs are obligatorily factored into the meaning of (71)-(73) (i.e., that we are here in presence of obligatory SIs, just like in the examples considered in section 4.1). It can be shown that SIs in (71)a, (72) and (73)a do not affect the DE character of the context, wheareas they do in (71b) and (73b). Thus, in these latter cases, we no longer have DE contexts once SIs are factored in and hence the condition for the proper licensing of NPIs is not met.

To see how SIs could affect downward entailingness, consider the relationship between the sentences in (74).

(74) a. John didn't talk both to students and professors
b. John didn't talk both to students and physics professors

(74a) entails (74b) as long as their potential scalar implicatures are not taken into account. But otherwise, this is no longer the case. The reason for this is that (74a) triggers the SI that the stronger alternative with disjunction – (75a) below–is false, namely the implicature that the speaker talked either to students or to professors. If we factor SIs into basic meanings (deriving *strengthened meanings*), (74a) no longer entails (74b): in a situation where John talked to two biology professors and to no one else, (74a) is true on its strengthened meaning but (74b), while true on its weak meaning, is false on its strengthened meaning. In other words, there is no entailment between the sentences in (74), if they receive the syntactic parse in (74)'.

(74)' a. O_{ALT}[John didn't talk both to students and to professors].
b. O_{ALT}[John didn't talk both to students and to physics professors].

The situation in (75) is very different. (75a) has no SIs, because the alternative with conjunction – (74a) – is a weaker alternative and is therefore not excluded by any of the approaches to SIs.

(75) a. John didn't talk to students or to professors.
b. John didn't talk to students or to physics professors.

So, even if (75) received a parse with *O*, parallel to (74)', this will not interfere with downward-entailingness:

(75)' a. O_{ALT}[John didn't talk both to students or to professors].
b. O_{ALT}[John didn't talk both to students or to physics professors].

O is vacuous in (75)', and therefore does not affect the entailment between the (a) and the (b) sentence. In other words, if *O* is inserted above negation, the NPI in (73b) is no longer in a downward entailing environment, while in (73b) downward entailment is not affected.

The same applies, arguably, to all interveners. For example, we can make sense of the fact that universal quantifiers are interveners but existential quantifiers are not. Existential quantifiers, in contrast to universal quantifiers, are the lowest members of their scale. Existential quantifiers and universal quantifiers form a Horn scale in which the universal is the logically strong member. Since, strength is reversed under downward entailing operators, universal quantifiers lead to matrix implicatures when embedded under such operators and existential quantifiers do not. The relevant implicatures, in turn, destroy downward entailingness, thus yielding the intervention effect.

But of course, none of this can work if SI are computed outside grammar. Under such an architecture, there is no reason why they should affect the licensing of NPIs. Moreover, Chierchia shows that his account can work only under very specific assumptions about the effects of SIs on syntactic intervention effects. If there is something to the account, SIs clearly must be computed within grammar.

4.4. Free choice

An utterance of the sentence in (76) is typically interpreted as a license to choose freely between two available options (the free choice inference, henceforth Free Choice).

(76) You are allowed to eat cake or ice cream.
There is at least one allowed world where you eat cake or ice cream.

More specifically, (76) licenses the two inference in (77).

(77) Free Choice (inference of ((76))
a. You are allowed to eat cake.
There is at least one allowed world where you eat cake.
b. You are allowed to eat ice cream.
There is at least one allowed world where you eat ice cream.

Free Choice, however, does not follow in any straightforward way from the basic meaning of the sentences. (76) – which contains two logical operators: the existential modal *allowed* and the disjunction *or* – should express the proposition that the disjunction holds in at least one of the allowed worlds [◊(C ∨IC)]. And, the truth of this proposition does not guarantee that for each disjunct there is an allowed world in which the disjunct is true. [◊(C ∨IC) ≠> ◊(C) ∧ ◊(IC).]

Kamp (1973), who identified the puzzle, suggested that it be resolved by strengthening the basic semantics of the construction and a solution along such lines has been worked out also in Zimmermann (2000) and Geurts (2005) (see also Simons 2005). However, Kratzer & Shimoyama (2002) – henceforth K&S – and Alonso-Ovalle (2005) pointed out that such a revision would get the wrong result when the construction is embedded in a downward entailing environment:

(78) No one is allowed to eat cake or ice cream

If (76) – as part of its basic meaning – were to entail Free Choice, we would expect (78) to be true if one of the free choice inferences in (77) were false for every individual in the domain (e.g., if half the people were allowed to eat cake the other half were allowed to eat ice cream, but no one was free to choose between the two desserts). But (78) seems to express a much stronger proposition, namely that no one is allowed to eat cake and that no one is allowed to eat ice cream.

We've already seen this pattern in section 4.1, namely an inference that appears when a sentence is uttered in isolation, but is not incorporated into its meaning when the sentence is further embedded in a downward entailing environment. We've also seen that this otherwise puzzling pattern would follow straightforwardly if the inference could be derived as an SI. In the case of Free Choice, K&S suggest that the inference should follow from a reasoning process about the belief state of the speaker that one might call meta-implicature (see also, e.g., Schulz (2005), Klinedinst (2006), Chemla (2008), for related proposals).

Specifically, K&S suggest that the sentences in (76) has the alternatives given in (79) below, for which we've argued on independent grounds in section 3.

(79) Alternatives for (76) proposed by K&S/Alonso-Ovalle:
   a. You are allowed to eat the cake.
   b. You are allowed to eat the ice cream.

Furthermore, they suggest that when a hearer h interprets s's utterance of (76), h needs to understand why s preferred (76) to the two alternatives. K&S, furthermore, suggest that it is reasonable for h to conclude that s did not choose the alternative because she was not happy with their strong meaning (basic meaning + implicatures) – hence our term meta-implicature. Specifically, K&S suggest that h would attribute s's choice to the belief that the strong meanings of (79a) and (79b) (stated in (80)) are both false.

(80) Strong meaning of the alternatives for (76)
   a. You are allowed to eat the cake and you are not allowed to eat the ice cream.
   b. You are allowed to eat the ice cream and you are allowed to eat the cake.

And, as the reader can verify, if (76) is true and the strengthened alternatives in (80) are both false, then the Free Choice inferences in (77) have to be true.

We believe this logic is basically correct, but we don't see a way to derive it from basic principles of communication (Maxims). In fact, if s believed that the Free Choice inferences hold, the Maxim of Quantity would have forced s to prefer the sentences in (79) and to avoid an utterance of (76) altogether. The fact that s did not avoid (76) should, therefore, lead h to the conclusion that s does not believe that the sentences in (79) are true (see our discussion in section 1.1.).

This has led Chierchia (2006) and Fox (2007) to provide a formal/grammatical alternative to K&S. We cannot go over the details of the proposals, but would like to point out Fox's observation, namely that K&S's results follow from a representation in which two instances of the operator, *O*, are appended to (76):

(81) Alternatives for A logical form for (76) that derives Free Choice:
O[O(You are allowed to eat cake or ice cream)].
There is at least one allowed world where you eat cake or ice cream.
And (80)a,b, are both false.

So while there are reasons to believe that free choice effects can be explained in a principled way as meta- (or higher order) implicatures (Spector 2007a, which we briefly discussed in subsection 4.1, also makes use of higher-order implicatures) and while the basic idea might seem at first compatible with a Gricean system, working it out turns out to be in conflict with the basic maxims. Furthermore, Chierchia (2006) further develops this logic arguing that constraints on the relevant grammatical representations yields an account of the cross-linguistic distribution of Free Choice items. If these proposals are anywhere close to the mark, then clearly implicatures must be part of grammar

4.5. Non-monotonic contexts: negating alternatives that are neither stronger nor weaker

Consider the following sentence:

(82) Exactly one student solved some of the problems

Let's assume that (82)'s only scalar alternative is (83).

(83) Exactly one student solved all of the problems

(83) is neither stronger nor weaker than (82): both sentences can be true in a situation where the other is false. Since (83) is not more informative than (82), Grice's maxim of quantity, under its most natural understanding, does not require that one utter (83) rather than (82) even when both are believed to be true and relevant. So the Gricean approach, unless supplemented with quite specific assumptions, predicts no SI in the case of (82). In contrast to this, a theory that incorporates the exhaustivity operator, which is modeled on the semantics of *only*, does predict an implicature for (82) when the exhaustivitiy operator takes matrix scope (it also predicts the possibility of an embedded implicature under the scope of *exactly one*, as discussed in section 2.2).

Indeed, applying the exhaustivity operator to a given sentence S with alternatives ALT(S) generally returns the conjunction of S and of the negations of all the alternatives of S that are not entailed by S (modulo the modification we have introduced above in order to reach a correct treatment of disjunctive sentences), which include both alternatives that are stronger than S and possibly alternatives that are neither stronger nor weaker than S. So the meaning of (82) (under the parse *O(Exactly one student solved some of the problems*)) is predicted to be the proposition expressed in (84a) which is equivalent to (84b) (Note that the fact that exactly one student solved at least one problem entails that no student except one solved any problem at all):

(84) a. Exactly one student solved at least one of the problems and it is false that exactly one student solved all of the problems

b. There is a student *x* who solved at least one of the problems, no student except *x* solved any problem, and *x* did not solve all of the problems.

It seems to us that this prediction is borne out: (84a), which is equivalent to (84b), is indeed a very natural interpretation for (82). The fact that implicature computation seems to involve the negation of non-stronger alternatives is quite unexpected from the Gricean perspective, unless further assumptions are made (for instance the assumption that the set of alternatives of any sentence is closed under conjunction, cf. Van Rooij & Schulz 2004, 2006 and Spector 2003, 2006, 2007b; see Fox 2007 for an argument against this assumption).

4.6. Constraints on the placement of the exhaustivity operator

We have observed that a hallmark of SIs is that they tend to disappear in downward-entailing environments – i.e., the strengthened reading of scalar items is dispreferred under, say, negation or in the restrictor of a universal quantifier. At first sight, this phenomenon makes the pragmatic, neo-Gricean, account of SIs particularly appealing: indeed, as we have seen, the absence of the strengthened reading in DE contexts is directly predicted by the neo-Gricean perspective.

However, as we pointed out in section 2, the strengthened meaning of a scalar item is actually not entirely ruled out in DE contexts; it is only dispreferred. From a purely Gricean perspective, it is a challenge to explain why a scalar item could ever be interpreted under its strengthened meaning in a DE context (so called 'intrusive implicatures'). To account for such cases, advocates of the purely pragmatic perspective are forced to introduce new mechanisms, such as a mechanism of metalinguistic negation, which then has to be generalized (cf. Horn 1989) to other operators (but if our previous arguments are conclusive, these 'repairs' are anyway unable to account for the full range of phenomena). The grammatical view does not face a similar challenge; but it clearly needs to be supplemented with some principles that determine which particular readings are preferred and which ones are dispreferred (and hence marked).

One possibility that suggests itself is that, when a sentence is potentially ambiguous, there is a preference for the strongest possible interpretation. Such a general principle has been suggested independently by various researchers beginning with Dalrymple et al. (1998) – the 'strongest meaning hypothesis'. If a principle of this sort is adopted, then inserting *O* in a DE context would be dispreferred: indeed, for any sentence S, O(S) is stronger than S; hence, inserting O(S) in the (immediate) scope of a DE operator X, i.e., an operator that reverses logical strength, gives rise to a sentence X(O(S)) that is now *weaker* than what would have resulted if O were absent, i.e., weaker than X(S).

How exactly such a principle should be stated is far from trivial. Given our general perspective, it should state that a Logical Form S in which *O* occur is dispreferred in case S is a-symmetrically entailed by some well-defined competitor S'. That is, we have to define, for any such S, its *comparison class*. One possibility we will briefly consider is the following: an occurrence of the exhaustivity operator is dispreferred if and only if it gives rise to a reading that is weaker than what would have resulted in its absence.

(85) Strongest Meaning Hypothesis
Let S be a sentence of the form [$_S$ . . . O(X) . . .]. Let S' be the sentence of the form [$_{S'}$ . . .X . . .], i.e., the one that is derived from S by replacing O(X) with X, i.e. by eliminating this particular occurrence of O. Then, everything else being equal, S' is preferred to S if S' is logically stronger than S.

(85) predicts that O should be dispreferred under DE-operators. However, it does not predict any preference between matrix exhaustification and embedded exhaustification in UE-contexts. This can be illustrated by the following example:

(86) For this class, we must read most of the books on the reading list

An exhaustivity operator could be inserted either above or below the modal *must*, giving rise to the following readings:

(87) a. O(we must read most of the books on the reading list)
= we must read most of the books on the reading list and we don't have to read all of then
b. We must O(read most of the books on the reading list)
= we must read most of the books on the reading list and we have the obligation no to read them all

(87b) a-symmetrically entails (87a). Yet the principle in (85) does not predict that (87b) should be preferred to (87a) (or the other way around). This is so because according to (85), (87a) and (87b) do not belong to each other's comparison classes. Rather, each of them is to be compared to the proposition that one gets be deleting the operator, namely to the non-strengthened reading of 'We must read *Ulysses* or *Madame Bovary*". Plainly, the condition stated in (85) is met in both cases, since in both cases the presence of O has a strengthening effect, and so no preference is predicted. More generally, in UE contexts, (85) does not favor one particular insertion site for the exhaustivity operator. Of course, more general considerations (such as, for instance, the plausibility of a given reading) might create a preference for certain readings.

5. Concluding remarks

In this article we tried to show that SIs can occur in all sorts of embedded context. If this attempt has been successful, we think it calls for a reassessment of the semantics/pragmatics interface. In order to establish our point, we have adopted the view that implicatures arise through a silent exhaustification operator, akin to *only*, which acts on scalar alternatives. We think that the idea – while leaving many open issues – has significant benefits: in many cases (involving Hurford's Constraint, iterated applications of *O*, etc.) it makes just the right predictions.

The grammatical view of SIs retains the most beautiful feature of the Gricean insight: the sensitivity of SIs to embeddings within polarity affecting contexts. And, through the

link to alternative sensitive operators, this view also creates a powerful bridge to a host of like phenomena occurring in very diverse corners of grammars (from the analysis of plurals, through free choice, to intervention and the like). Within the limits of the present article, these remain largely promissory notes. But we hope that we were able to lay out the strategy that needs to be pursued in a fairly clear manner. Finally, we hope that it will be possible begin to reap additional benefits from the entrance of SIs (and of, possibly, implicatures of other sorts) into the computational system of grammar.

We would like to thank participants in various classes, colloquia and seminars in which this work was presented (by at least one of us), in particular Columbia University (2008), Institut Jean-Nicod (2008), the LSA summer institute (2005), University of Georgetown (2006), LOGOS (Barcelona 2008), Harvard-MIT seminar in semantics (2006), MIT-Paris workshop in linguistics (2007), and the University of Vienna (2009). Special thanks to Emmanuel Chemla, Irene Heim, Maribel Romero, Daniel Rothschild and Philippe Schlenker.

6. References

Alonso-Ovalle, Luis 2005. Distributing the disjuncts over the modal space. In: L. Baterman & C. Ussery (eds.). *Proceedings of the North Eastern Linguistic Society (= NELS) 35*. Amherst, MA: GLSA, 75–86.

Atlas, Jay D. & Stephen Levinson 1981. *It*-clefts, informativeness, and logical form: Radical pragmatics (revised standard version). In: P. Cole (ed.). *Radical Pragmatics*. New York: Academic Press, 1–61.

Breheny, Richard 2008. A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics* 25, 93–140.

Carnap, Rudolf 1950. *Logical Foundations of Probability*. Chicago, IL: The University of Chicago Press.

Carston, Robyn 1988. Implicature, explicature, and truth-theoretic semantics. In: R. Kempson (ed.). *Mental Representations: The Interface Between Language and Reality*. Cambridge: Cambridge University Press, 155–181.

Chemla, Emmanuel 2008. *Présuppositions et implicatures scalaires. Etudes formelles et expérimentales*. Ph.D. dissertation. Ecole des Hautes Etudes en Sciences Sociales, Paris.

Chierchia, Gennaro 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In: A. Belletti (ed.). *Structures and Beyond 3*. Oxford: Oxford University Press, 39–103.

Chierchia, Gennaro 2006. Broaden your views. Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry* 37, 535–590.

Chierchia, Gennaro, Danny Fox & Benjamin Spector 2009. Hurford's constraint and the theory of scalar implicatures. In: P. Egré & G. Magri (eds.). *Presuppositions and Implicatures. Proceedings of the MIT-Paris Workshop* (MIT Working Papers in Linguistics 60). Cambridge, MA: MIT, 47–62.

Dalrymple, Mary, Makoto Kanazawa, Yookyung Kim, Sam Mchombo & Paul S. Peters 1998. Reciprocal expressions and the concept of reciprocity. *Linguistics & Philosophy* 21, 159–210.

Davis, Wayne 1998. *Implicature: Intention, Convention, and Principle in the Failure of Gricean Theory.* Cambridge: Cambridge University Press.

Davis, Wayne 2005. Implicature. In: E.N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy (Summer 2005 Edition)*. http://plato.stanford.edu/entries/implicature/, February 26, 2012.

Ducrot, Oswald 1973. *La preuve et le dire*. Paris: Mame.

Fauconnier, Gilles 1975a. Polarity and the scale principle. In: R. Grossman, L. San & T. Vance (eds.). *Papers form the Eleventh Regional Meeting of the Chicago Linguistic Society (= CLS)*. Chicago, IL: Chicago Linguistic Society, 188–199.

Fauconnier, Gilles 1975b. Pragmatic scales and logical structure. *Linguistic Inquiry* 6, 353–376.

Fox, Danny 2004. *Implicatures and Exhaustivity.* Lecture notes (Class 4: Back to the Theory of Implicatures, University of Southern California, November 2004). http://web.mit.edu/linguistics/people/faculty/fox/class_4.pdf, February 26, 2012.

Fox, Danny 2007. Free choice disjunction and the theory of scalar implicatures. In: U. Sauerland & P. Stateva (eds.). *Presupposition and Implicature in Compositional Semantics*. New York: Palgrave Macmillan, 71–120.

Fox, Danny & Martin Hackl 2006. The universal density of measurement. *Linguistics & Philosophy* 29, 537–586.

Fox, Danny & Roni Katzir 2009. On the characterization of alternatives for implicature and for focus. In: P. Egré & G. Magri (eds.). *Presuppositions and Implicatures. Proceedings of the MIT-Paris Workshop* (= MIT Working Papers in Linguistics 60). Cambridge, MA: MIT, 101–110.

Gamut, L. T. F. 1991. *Logic, Language, and Meaning*. Chicago, IL: The University of Chicago Press.

Gazdar, Gerald 1979. *Pragmatics: Implicature, Presupposition and Logical Form*. New York: Academic Press.

Geurts, Bart 2005. Entertaining alternatives: Disjunctions as modals. *Natural Language Semantics* 13, 383–410.

Geurts, Bart 2009. Scalar implicature and local pragmatics. *Mind & Language* 24, 51–79.

Grice, Paul 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Groenendijk, Jeroen & Martin Stokhof 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph.D. dissertation. University of Amsterdam.

Groenendijk, Jeroen & Martin Stokhof 1990. *Partitioning Logical Space.* Annotated handout for the *European Summer School in Language, Logic and Computation (= ESSLI)*, Leuven, August 1990. http://staff.science.uva.nl/~stokhof/pls.pdf, February 26, 2012.

Guerzoni, Elena 2000. Towards a movement-based account of the locality constraints on negative polarity items. In: C. Czinglar et al. (eds.). *Proceedings of the Conference of the Student Organisation of Linguistics in Europe (= CONSOLE) 8*, Leiden: SOLE, Leiden University, 125–138.

Guerzoni, Elena 2004. Even-{NPI}s in yes/no questions. *Natural Language Semantics* 12, 319–343.

Guerzoni, Elena 2006. Intervention effects on NPIs and feature movement: Towards a unified account of intervention. *Natural Language Semantics* 14, 359–398.

Heim, Irene 1991. Artikel und Definitheit. In: A. v. Stechow & D. Wunderlich (eds.). *Semantik – Semantics. Ein internationales Handbuch der zeitgenössischen Forschung – An International Handbook of Contemporary Research* (HSK 6). Berlin: de Gruyter, 487–535.

Hirschberg, Julia 1985. *A Theory of Scalar Implicature*. Ph.D. dissertation. University of Pennsylvania, Philadelphia, PA.

Horn, Larry 1972. *On the Semantic Properties of Logical Operators in English*. Ph.D. dissertation. University of California, Los Angeles, CA.

Horn, Larry 1985. Metalinguistic negation and pragmatic ambiguity. *Language* 61, 121–174.

Horn, Larry 1989. *A Natural History of Negation*. Chicago, IL: The University of Chicago Press.

Horn, Larry 2006. The border wars: A neo-Gricean perspective. In: T. Turner & K. von Heusinger (eds.). *Where Semantics meets Pragmatics*. Amsterdam: Elsevier, 21–48.

Hurford, James 1974. Exclusive or inclusive disjunction. *Foundation of Language* 11, 409–411.

Kamp, Hans 1973. Free choice permission. *Proceedings of the Aristotelian Society* 74, 57–74.

Katzir, Roni 2007. Structurally-defined alternatives. *Linguistics & Philosophy* 30, 669–690.

Klinedinst, Nathan 2006. *Plurality and Possibility*. Ph.D. dissertation. University of California, Los Angeles, CA.

Kratzer, Angelika & Junko Shimoyama 2002. Indeterminate pronouns: The view from Japanese. In: Y. Otsu (ed.). *Proceedings of the Tokyo Conference on Psycholinguistics 3*. Tokyo: Hituzi Syobo, 1–25.

Krifka, Manfred 1993. Focus and presupposition in dynamic interpretation. *Journal of Semantics* 10, 269–300.

Krifka, Manfred 1995. The semantics and pragmatics of polarity items. *Linguistic Analysis* 25, 209–257.

Kroch, Anthony 1972. Lexical and inferred meanings for some time adverbs. *Quarterly Progress Reports of the Research Laboratory of Electronics* 104, 260–267.

van Kuppevelt, Jan 1996. Inferring from topics: Scalar implicatures as topic dependent inferences. *Linguistics & Philosophy* 19, 393–443.

Ladusaw, Bill 1979. *Polarity Sensitivity as Inherent Scope Relations*. Ph.D. dissertation. University of Texas, Austin, TX.

Landman, Fred 1998. Plurals and maximalization. In: S. Rothstein (ed.). *Events and Grammar*. Dordrecht: Kluwer, 237–272.

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, Stephen C. 2000. *Presumptive Meaning*. Cambridge, MA: The MIT Press.

Lewis, David 1973. *Counterfactuals*. Oxford: Blackwell.

Linebarger, Marcia 1987. Negative polarity and grammatical representation. *Linguistics & Philosophy* 10, 325–387.

Magri, Giorgio 2009. A theory of individual level predicates based on blind scalar implicatures. *Natural Language Semantics* 17, 245–297.

Matsumoto, Yo 1995. The conversational condition on Horn scales. *Linguistics & Philosophy* 18, 21–60.

Meier, Marie-Christine & Uli Sauerland 2008. A pragmatic constraint on ambiguity detection. *Natural Language and Linguistic Theory* 27, 139–150.

Récanati, François 2003. Embedded implicatures. *Philosophical Perspectives* 17, 1299–1332.

van Rooij, Robert & Kathrin Schulz 2004. Exhaustive interpretation of complex sentences. *Journal of Logic, Language and Information* 13, 491–519.

van Rooij, Robert & Kathrin Schulz 2006. Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics & Philosophy* 29, 205–250.

Rooth, Mats 1985. *Association with Focus*. Ph.D. dissertation. University of Massachusetts, Amherst, MA.

Rooth, Mats 1992. A theory of focus interpretation. *Natural Language Semantics* 1, 117–121.

Russell, Benjamin 2006. Against grammatical computation of scalar implicatures. *Journal of Semantics* 23, 361–382.

Sauerland, Uli 2003. A new semantics for number. In: R. Young & Y. Zhou (eds.). *Proceedings of Semantics and Linguistic Theory (= SALT) XIII*, Ithaca, NY: Cornell University, 258–275.

Sauerland, Uli 2004. Scalar implicatures in complex sentences. *Linguistics & Philosophy* 27, 367–391.

Schulz, Kathrin 2005. A pragmatic solution for the paradox of free choice permission. *Synthese* 147, 343–377.

Simons, Mandy 2005. Dividing things up: The semantics of *or* and the modal/*or* interaction. *Natural Language Semantics* 13, 271–316.

Singh, Raj 2008. On the interpretation of disjunction: Asymmetric, incremental, and eager for inconsistency. *Linguistics & Philosophy* 31, 245–260.

Soames, Scott 1982. How presuppositions are inherited: A solution to the projection problem. *Linguistic Inquiry* 13, 483–545.

Spector, Benjamin 2003. Scalar implicatures: Exhaustivity and Gricean reasoning. In: B. ten Cate (ed.). *Proceedings of the European Summer School in Logic, Language and Information Student Session (= ESSLLI Student Session) 8*. Vienna, Austria, 277–288. http://users.soe.ucsc.edu/~btencate/esslli03/stus2003proc.pdf, May 8, 2011.

Spector, Benjamin 2006. *Aspects de la pragmatique des opérateurs logiques*. Ph.D. dissertation. Université Paris 7.

Spector, Benjamin 2007a. Aspects of the pragmatics of plural morphology. In: U. Sauerland & P. Stateva (eds.). *Presupposition and Implicature in Compositional Semantics*. New York: Palgrave Macmillan, 243–281.

Spector, Benjamin 2007b. Scalar implicatures: Exhaustivity and Gricean reasoning. In: M. Aloni, A. Butler & P. Dekker (eds.). *Questions in Dynamic Semantics*. Amsterdam: Elsevier, 225–249.

Sperber, Dan & Deirdre Wilson 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.

Stalnaker, Robert 1968. A theory of conditionals. In: N. Rescher (ed.). *Studies in Logical Theory.* Oxford: Blackwell, 98–112.

Zimmermann, Thomas E. 2000. Free choice disjunction and epistemic possibility. *Natural Language Semantics* 8, 255–290.

Gennaro Chierchia, Cambridge, MA (USA)
Danny Fox, Cambridge, MA (USA)
Benjamin Spector, Paris (France)