Between "Cost" and "Default" of Scalar Implicature: the Cost of Embedding*

Francesca Foppolo
Dept. of Psychology
University of Milano-Bicocca

francesca.foppolo@unimib.it

Abstract

Two main points constitute a matter of debate concerning the phenomenon of Scalar Implicatures (SIs): the *place* of their derivation, which opposes a "recursive"/grammatically driven approach such as Chierchia's (Chierchia, 2002&2006; Fox, 2003; Landman, 1998; Levinson, 2000) to traditional Neo-Gricean approaches that view SIs as genuinely post-grammatical/pragmatic processes that are added "globally", independently of compositional semantics (Russell, 2006; Sauerland, 2005; Spector, 2003 a.o.); and the question of the processing *cost* of SI computation, which most of the experimental works on SIs have recently been focused on (Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2005; Noveck & Posada, 2003). Orthogonal to this debate, our contribution is based on the assumption that SIs are derived locally (following Chierchia, 2006) and tests the effect of logical abstract properties of the context (e.g. monotonicity) on the computation of implicatures and their cost. Our main finding is that a "cost" is found only when implicatures are added despite the fact that they lead to a weakening of the overall assertion (namely, in Downward Entailing contexts): this loss in informativity, and not implicature computation *per se*, is interpreted as the source of this "cost".

*I would like to thank Gennaro Chierchia, Maria Teresa Guasti, Francesca Panzeri and Maria Nella Carminati for their encouragment, help and discussion during the preparation of this work.

Grønn, Atle (ed.): *Proceedings of SuB12*, Oslo: ILOS 2008 (ISBN 978-82-92800-00-3), 137–150.

1 Introduction

Consider the example in (1):

(1) Lorenzo is singing *or* dancing

This sentence normally conveys the fact that its (more informative) alternative (2) doesn't hold. Hence (3), which is how (1) is normally interpreted:

(2) Lorenzo is singing *and* dancing

(3) Lorenzo is singing *or* dancing but not both

The Scalar Implicature added in (3) corresponds to the *exclusive* interpretation of "or" (A or B but not both), which contrasts with the *inclusive*/logical meaning of "or" (A or B or both). The mechanism by which SIs are derived is based on the notion of scale on the one hand, and on that of informational strength on the other (Grice, 1957; Horn, 1972). By virtue of the fact that (2) contains the stronger element "and" (given the scale <or, and>, where "and" entails "or"), and that (2) is not what was actually said, then one is entitled to assume that (2) does not hold, hence (3), in which the negation of the strongest alternative is added.

The main question that has been investigated experimentally so far is whether implicatures in *unembedded* or *root* contexts are costly or not. In this perspective, most of the experimental work aimed at measuring the processing cost of intepreting (1) as (3) in order to find evidence in support of one of two opposite theoretical approaches to SIs: Default approaches on the one hand (a.o. Chierchia, 2004; Levinson, 2000), that treat implicature computation as something that our computational/processing system performs automatically to maximise information content thus, by definition, is virtually costless (a claim which is in fact also shared by most Neo-Gricean approaches to SI); and Context-Driven theories on the other, like Relevance Theory (Sperber & Wilson, 1986), according to which every operation imposed to our processing system must be evaluated in terms of "costs and benefits", ultimately in terms of "relevance" to contextual assumptions so that only those stimuli that are relevant enough are worth a processing effort. As we shall see, our main point is orthogonal to the issue of a general *cost* of SI derivation and bears on it only indirectly.

In a broader theoretical perspective, the question of SIs is central to the ongoing debates concerning the definition of the status and interfaces of syntax, semantics and pragmatics. In this respect, different accounts have been developed to explain how and when implicatures are derived. Entering the details of this debate, however, goes well beyond the purposes of this work, whose main aim is that of presenting the results of a novel experimental work on SIs. The theoretical background of this work is constituted

by a recent paper by Chierchia (Chierchia, 2006), in which he proposes a unified account of Negative Polarity (NPIs), Free Choice (FCIs) and Scalar Items, building on the notion of "domain widening", polarity sensitivity and a general principle of pragmatic strengthening (cf. also Krifka, 1996). Specifically, for what concerns the phenomenon under discussion here, his main claim is that certain "pragmatic" processes, such as the process of deriving SIs, are part of the recursive/computational system: a binary feature σ is introduced as regulating the activation of scalar alternatives associated to scalar and polarity sensitive items (PSIs). This feature can be assigned two values: [±σ]. Selecting [+σ] results in the activation of the scalar/domain alternatives; selecting [-σ] results in the selection of the plain meaning in which the alternatives are not considered. While NPIs and FCIs obligatorily activate domain alternatives (i.e., always select [+σ] to be grammatical), Scalar Items only *optionally* activate their scalar alternatives. Once they are activated, they are factored into meaning via an alternative sensitive operator **O** similar to *Only* (cf. Fox, 2003). I won't pursue further the discussion on the parallelism of scalar and PSIs (this goes beyond the purposes of the present paper) but it's interesting to report a generalization on SIs already reported in Chierchia, 2002 (see also Kadmon & Landman, 1993): "(Ordinary) scalar implicatures are suspended in the contexts that license *any* (as a Neg Pol or as Free Choice Item)". Typically, these are the contexts defined as Downward Entailing (DE, or Downward Monotone), i.e. those contexts that licence inferences from sets to their subsets. For example, the antecedent of conditional represents a DE context, in contrast with the consequent of conditional, which represents an Upward Entailing context instead, allowing only inferences from a set to its superset. Crucially, adding an implicature in DE contexts leads to a weakening of the overall assertion (given that informativity is "reversed" in DE contexts), while it leads to a strengthening in case the scalar term appears in a NON-DE context. Considering our general tendency to be maximally informative and the monotonicity properties of the context, this is in fact how we normally interpret sentences like (4), representing a DE context, and (5), representing a NON-DE context.

(4) <u>If Lorenzo is singing or dancing</u> (*or both*) he's happy [DE]

(5) If Lorenzo is happy, <u>he is singing or dancing</u> (*not both*) [NON-DE]

Let's assume that scalar alternatives are activated in the examples above, i.e. that disjunction is interpreted as $or_{[+\sigma]}$. In terms of Chierchia's recursive approach, this would lead to the following interpretations, in which the SI is computed *locally* as soon as the scalar trigger is encountered:

(4') [if **O** (Lorenzo is singing or dancing), then he is happy] [DE]
=if Lorenzo is singing or dancing <u>but not both</u>, then he is happy

(5') [if Lorenzo is happy, then **O** (he is singing or dancing)] [NON-DE]
=If Lorenzo is happy, then he is singing *or* dancing <u>but not both</u>

Consider (4'): adding the SI locally, thus selecting the *exclusive* meaning of "or", would allow the inference in (4''), that really seems odd and unwarranted:

(4'') If Lorenzo is singing and dancing, then he is *not* happy

Taking into account the monotonicity of the context and our tendency to be maximally informative, the following distributional generalizations can thus be predicted:

i The exhaustive interpretation (via application of the operator **O)** of a scalar term is easier in NON-DE than in DE contexts, because it strengthens the assertion;
→ (5') is easier than (4')
ii. Having an implicature embedded in DE contexts is way harder than having it embedded in NON-DE contexts, because it weakens the assertion
→ (4') is harder than (5')
iii. The flip between having an implicature and not having it is relatively easy in NON-DE contexts
→ the activation of scalar alternatives is *optional* in case of scalar items
iv The flip between having an implicature and not having it is hard in DE contexts

2 The cost of embedding

In this section I will present an experimental study that tested the distributional generalization listed above with respect to the interpretation of *or* embedded in contexts that differ in monotonicity.

2.1 Participants

Thirty subjects participated in this experiment. They were mainly students at the Psychological Faculty of the University of Milano-Bicocca, and received credits for their participation.

2.2 Material and procedure

Participants were tested individually in a quiet room using a laptop. Their task was to evaluate sentences in certain situations, judging them "true" or "false" with respect to a scenario consisting of a block of four pictures to be considered as a whole. They were also told to be "charitable": whenever they encountered a sentence that could bear more than one interpretation, they should choose the one that rendered the sentence true, even if that interpretation was not their favoured one. To familiarize them with the procedure, they were shown a training session in which they were assisted by the experimenter. During this training, they encountered sentences that were clearly true in the scenario,

some that were clearly false and some that were somehow ambiguous. For example, they were presented with the sentence "Two girls are sitting on a chair" in a scenario depicting four different girls, each of them sitting on a chair. Typically, many subjects would tend to judge the sentence false in such a scenario, interpreting "two" as "exactly two" instead of being "charitable" and accessing the logical "at least two" interpretation. Whenever this happened, the experimenter prompted the participant to be charitable and ask her to revise her interpretation of the sentence accordingly as to make it true, if she found a way to do it. To keep track of this operation, participants were also asked the following question whenever they answered "true": "How much do you think the sentence is a good description of the situation represented in the pictures?" They were given a scale of response varying from 1 (bad) to 5 (good). In cases like the example above, we expected subjects that were prompted to be charitable to select a low score on the scale despite the fact that they accepted the sentence in the end.

The experiment proceeded as follows: each single sentence appeared in white at the top of a black screen. By pressing the space bar, a scenario consisting of four pictures appeared below the sentence. It's important to remember that the four pictures appeared altogether on the screen and were to be considered as a whole, unique scenario, representing the whole world to be taken into account in order to judge the sentence. Participants had to evaluate the sentence in such a scenario, pressing a green key if they judged it "true" and a red key if they judged it "false". Time taken to make a decision was recorded, starting from the moment they pressed the key to make the pictures appearing on the screen, till they pressed the answer key. Each subject was shown the complete battery of the material but saw only one occurrence per each critical item type, for a total of 17 test items, 4 of which were critical test sentences containing "or", and the others were controls and fillers. To avoid interferences from extra-linguistic factors, we only used fantasy names in the sentences during the experimental session. After the training session, subjects were told that they would explore different situations in planets different from Earth, meeting alien characters that used objects that are unfamiliar to inhabitants of Earth. They were also reassured that they were not required to memorize the names of these characters and objects, given that they would be provided with a description of each unfamiliar object immediately before each trial. Below, I provide an example of an introductory screen used before one test trial:

Fig. 1: Introductory screen: an example

The main purpose of having unfamiliar objects was that of ruling out world knowledge as much as possible: it's a well known fact that our expectations about how things go in usual circumstances may affect the computation or suspension of SIs. Think, for instance, at standard examples like "If you take the soup of the day or curry rice you'll pay the special price of 10$": solely guided by our world knowledge (and independently of the monotonicity of the context!), we would never expect to pay 10$ for taking both. However, given that our world knowledge do not extend to planet Glimp, we should bear no expectations whatsoever about situations that involve curps or dorfs or combinations of the two.

The experiment presented a 2×2 critical condition within subject design. First of all, two different types of sentences containing "or" were presented, differing in monotonicity (DE vs. NON-DE), as exemplified by the following examples (remember that only fantasy names were used in the experimental session):

(6) Condition I: context monotonicity
a. If a Glimp has a curp *or* a dorf, he also has a pencil [DE]
b. If a Glimp has a pencil, he also has a curp *or* a dorf [NON-DE]

Each sentence was presented in two different types of situation:

(7) Condition II: situations
S1 a situation in which the sentence is true on both readings
(I will refer to this as the "non-differentiating true" situation = NDT)
S2 a situation in which the sentence is true on one interpretation -crucially, the less informative one- but false on the other
(I will refer to this as the "differentiating-critical" situation = DC)

To well understand the experimental design, it's important to keep in mind that the two alternative interpretations of "or" are not logically *in*dependent of one another, given that one always entails the other and that the direction of entailment crucially depends on the monotonicity of the context: in NON-DE contexts, or_{exc} entails or_{inc}, thus the *exclusive* interpretation of "or" is the most informative in case of (b)-sentences; on the contrary, the *inclusive* interpretation of "or" is the most informative in case of (a)-

sentences, given that in DE contexts or_{inc} entails or_{exc}. Considering these entailment patterns, the critical conditions for each sentence types were, for instance, the ones reported below (presented to different subjects):

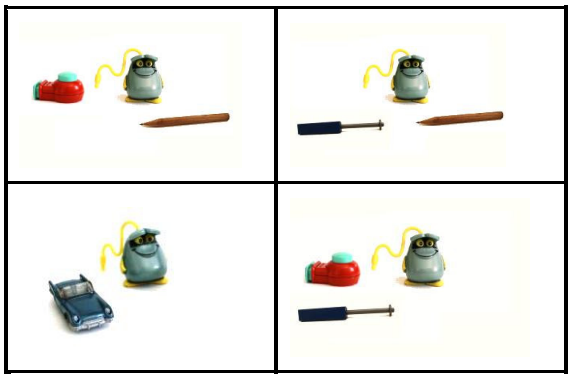


Fig. 2: (S1)
DC for DE context: *exclusive* "or" true (curp & dorf but no pencil)

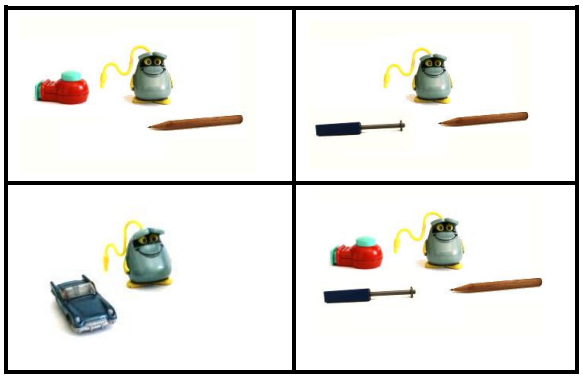


Fig. 3: (S2)
DC for NON-DE context: *inclusive* "or" true (curp & dorf but no pencil)

Note that the only crucial difference between the two scenarios is represented by the last picture in the sequence (during the experiment, the order of the pictures was randomized). Please note that these same configurations were also used as NDT conditions: for example, configuration S1 constituted the NDT condition for sentences of type (b) (i.e. NON-DE contexts) while S2 constituted the NDT condition for sentences of type (a) (i.e. DE contexts). Also, a control condition was added, that made the sentence false on any interpretation of "or" (non-differentiating false condition = NDF). For example, Fig. 4 was used as a control for (a) sentences:

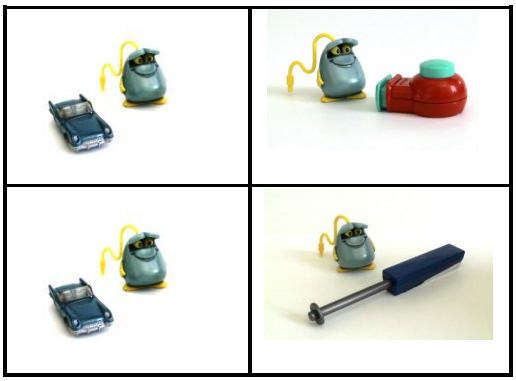


Fig. 4: (S3): control condition (NDF)

2.3 Results and Discussion

Results are summarized in Table 1 below[1]: column 4 reports the percentage of "true" answers; column 5 the rate assigned on the scale; columns 6-8 report respectively: the response times (RTs, in ms.) to answer "true" and "false" and the mean total time per condition (in parentheses the number of cases included in the analysis is shown).[2]

1	2	3	4	5	6	7	8
Sentence	**Context**	**Situation**	**True**	**Scale rate**	**RTs for True**	**RTs for False**	**Mean RTs (n.)**
(a)	**DE**	S1 (DC)	**57%**	3.47	11320	7167	9628 (n.27)
		S2 (NDT)	**90%**	3.81	8937	12362	9291 (n.30)
(b)	**NON-DE**	S2 (DC)	**77%**	4.04	10183	11754	10562 (n.29)
		S1 (NDT)	**87%**	4.38	9734	8341	9549 (n.29)

Table 1

By simply looking at the chart, one can immediately detect an interesting discrepancy between subjects' behavior on condition (a)-S1 (corresponding to the first row on the chart) and all the other conditions, a discrepancy that extends to all the measurements taken: it is the only condition in which subjects split; it gets the lowest rate on the scale; it takes the highest time to be accepted but the lowest time to be rejected.

Statistical analysis, in which different parameters were considered, gave support to this observation. First of all, I submitted our data to a 2 (context monotonicity: DE vs. NON-DE) x 2 (situations: S1 vs. S2) analysis of variance ANOVA using the proportion of "Yes" responses as the dependent measure. No significant effect of context monotonicity ($F(1, 116)=1.2787$, $p=.26048$) or situation ($F(1, 116)=2.5062$, $p=.11612$) was found, but a significant interaction of the two ($F(1, 116)=8.6437$, $p<.05$). Post-hoc analysis, by means of Fisher's LSD test, suggests that this effect is mainly due to a difference of the rate of acceptance of sentences of type (a) in condition S1 with respect to all the other conditions, as summarized as follows. Firstly, the proportion of subjects that accept (a)-sentences in condition S1 (corresponding to the DC, only *exclusive* condition for DE contexts) is significantly lower than the proportion of those that accept the same sentence in condition S2 (corresponding to the NDT condition) (57% vs. 90%, $p<.01$). Secondly, the proportion of acceptance of (a)-sentences in the critical condition

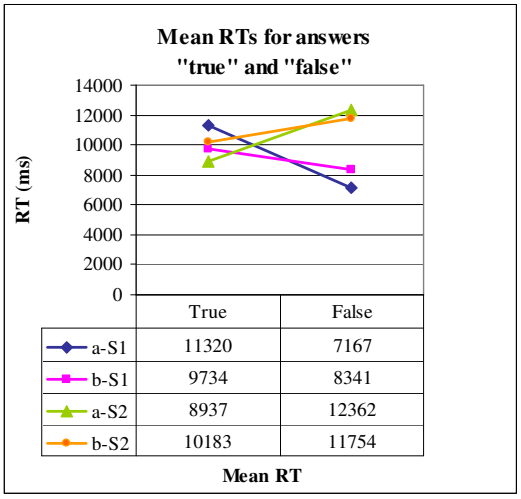
[1]In the chart, I don't report results on controls: consider that correct responses on these items are attested around 95% overall.

[2]Anomalous effects on RTs were curtailed in two steps: first, we excluded RTs exceeding 2.5 times the mean item time; then, values above individual cut-off (mean + 2 SD) were smoothed (over the total, 3,33% items were excluded and 3,75% smoothed).

S1 is also significantly different from the proportion of acceptance of (b)-sentences in the same condition (corresponding to the NDT condition for NON-DE contexts) (57% vs. 87%, p<.01). Thirdly, it is also marginally different from the proportion of acceptance of (b)-sentences in condition S2 (corresponding to the DC, only *inclusive* condition for NON-DE contexts) (57% vs. 77%, p=.057).[3] For what concern this last comparison, I would reasonably expect this discrepancy to be increased by removing the instruction "be charitable", that would have the effect of flattening the acceptance rate of sentences (a) in condition S1 (*exclusive* condition for DE contexts). Taken as a whole, these findings seem to suggest that adding an implicature in a DE context (i.e., accepting sentence (a) in the critical condition S1) is not a natural option that we would select automatically in our ordinary conversation, unless we are prompted to do so (as in this case, where participants were asked to be charitable). This same conclusion seem to arise from the comparisons across the rates assigned on the scale in case of "Yes" responses. Taking this rate as the dependent measure, I conducted an analogous 2x2 analysis of variance ANOVA, finding a significant effect of context (F(1,89)=5.866, p<.01). Fisher's LSD post-hoc test revealed that this effect is only due to the difference between the rate assigned on the scale when accepting sentence (a) in S1 (namely, 3.47) and the higher rate assigned when accepting sentence (b) in the same condition (namely, 4.38) (p<.01), a result that still seems to indicate that participants are less prone to accept *exclusive* "or" interpretation in DE than in NON-DE contexts.

Time taken to evaluate sentences was also analysed. A first point worthy of remark is the fact that no significant difference emerged in a 2x2 analysis of variance ANOVA taking context (DE vs. NON-DE, (F(1, 111)=.33, p=.57) or situation (S1 vs. S2, F(1, 111)=.11, p=.74) as critical factors, nor an interaction between the two (F(1, 111)=.42, p=.52). These results seem to indicate that the processing load required to evaluate sentences in both conditions was almost identical, at least considering mean RT. However, one needs to integrate this finding by considering the type of answer given ("True" vs. "False") separately, as plotted in the graph below:

[3] Conforming to the most standard procedure in the literature, I decided to perform an ANOVA to analyse my data (despite the presence of dichotomic variables). Note, however, that these same effects were replicated by means of an analysis of proportion: a-S1 vs. a-S2: $\chi^2(1, 60)=8{,}52$, p<.01; a-S1 vs. b-S1: $\chi^2(1, 60)=6.65$, p<.01.



Graph 1

Disregarding Condition I, thus independently of context monotonicity, a significant interaction between situation and type of answer was revealed by an ANOVA on RTs ($F(1, 111)=4,44$, $p<.05$). Post hoc analysis, by means of Fisher's LSD, revealed that this difference is due to the fact that the time taken to answer "False" in S2 is significantly higher than the time taken to answer "False" in S1 ($p<.05$). Interestingly, no other difference emerged. In particular, no significant difference was revealed between S1 and S2 considering the time taken to accept the sentences, independently of context monotonicity. Such a comparison, in fact, would be a crucial one to detect a presumptive "cost" of SI computation: in order to accept the sentences in scenario S1, that is compatible with *exclusive* "or", one should add the implicature, thus adding a "cost" to the base sentence processing time. Other pairwise comparisons between RTs were made, always taking the type of answer ("True" vs. "False") as the critical factor. Interestingly, only one comparison revealed statistically significant. Precisely, this was the time to accept (a)-sentences in S1 (DC – only *exclusive*- condition for DE contexts), compared to the mean time to reject it in the same condition ($t(25)=2.21$, $p<.05$). No other significant contrast emerged in analogous pairwise comparisons, and this fact is particularly intriguing if we take into account the predictions that non-Default theories would put forward in this case. In particular, Context-Driven theories would predict a difference between the time taken to accept (a) sentences in S1 (DC-only *exclusive*-condition for DE contexts) and the time taken to accept (b) sentences in S2 (DC-only *inclusive*- condition for NON-DE contexts), given that solely in the first situation an implicature must be added to accept the sentence. Also, they would predict higher RTs in rejecting than accepting (b) sentences in S2 (DC-only *inclusive*- condition for NON-DE contexts), given that such a rejection would be the effect of the addition of the SI associated to "or". According to their theoretical claims, the "cost" of deriving the implicature should result in increased processing time. Crucially, both the comparisons mentioned are far from being significant ($p=.52$ and $p=.61$ respectively). On the contrary, RT measures clearly seem to indicate that only subjects that accessed the

exclusive "or" interpretation in DE contexts did it at a "cost" (i.e., they took significantly longer to accept than to reject sentence (a) in S1, thus keeping with the more informative- *inclusive*- interpretation of "or"). Crucially, however, this "cost" seems should not be evaluated as a general addition to the processing load due to SI derivation *per se*, given that, as we mentioned, this "cost" does neither exceed the processing load required to accept (b) sentences in S2 (DC-only *inclusive*- condition for NON-DE contexts), nor is recorded when rejecting (b) sentences in such a situation.

All in all, our findings seem to conform to the distributional generalizations listed in section 1, and summarized below for convenience. In the first place, the predictions made in (i) and (ii) (i.e., (i): the exhaustive interpretation (via application of the operator **O**) of a scalar term seems easier in a NON-DE than in a DE context; (ii) having an implicature embedded in a DE context is way harder than having it embedded in a NON-DE context) are attested by the differences observed between sentences (a) (=DE context) and (b) (= NON-DE context) in situation S1 (DC –only *exclusive*- condition for DE contexts) in the rate of acceptance (57% vs. 87%, $p<.01$) and scale rate (3.47 vs. 4.38, $p<.01$). Secondly, the prediction in (iii) (i.e.: the flip between having an implicature and not having it is relatively easy in NON-DE contexts) is attested by the fact that participants treated sentences (b) alike in the two situations (in this case, the differences recorded in the rate of acceptance and in the RTs are only numerical, not statistically significant). Also, the fact that 77% of participants accepted (b)-sentences in situation S2 (DC –only *inclusive*- condition for NON-DE contexts) may reflect a "charitable" strategy, ultimately it may be evaluated as the effect of the instruction given. In terms of Chierchia's analysis, it may reflect the choice of selecting $or_{\text{-}\sigma}$, leaving the scalar alternatives inactive, to conform to the strategy suggested by the experimental setting. Lastly and most importantly, prediction (iv) (i.e.: the flip between having an implicature and not having it is hard in DE contexts) is attested by the major findings obtained in case of sentence (a) in S1 (DC –only *exclusive*- condition for DE contexts) and discussed above in details: namely, the fact that this condition got the lowest acceptance rate, the lowest rate on the scale and the highest RTs in case of acceptance.

3 Concluding remarks

One of the questions addressed in this experiment was the influence of context, ultimately the role of monotonicity and its effect on informativity, on SI computation. Taking into consideration subjects' distribution in accepting/rejecting the critical sentences in the relevant conditions, it seems that our results provide a clear answer to this question. In the first place, subjects clearly treat the two sentences differently. In particular, they derive SIs more when "or" appears in NON-DE than in DE contexts: in a situation compatible with the *exclusive* reading of "or", like S1, they accept sentences (b) significantly more than (a). This distribution is a hint that subjects are sensitive to

abstract logical properties, such as monotonicity, when evaluating sentences containing scalar items. Secondly, they treat (a)-sentences (representing DE contexts) differently in the two situations: namely, they overwhelmingly accept them in a situation compatible with *inclusive* "or" (S2), but they split in a situation in which only *exclusive* "or" interpretation makes the sentence true (S1). Again, this result is an indication that subjects are aware of informativity, which crucially links to monotonicity: computing SIs in DE contexts weakens the overall assertion, and this may be the reason why accepting (a) sentences in S1 is a less likely and a more unfelicitous option (as revealed by the lowest acceptance rate and scale rate) and it is more costly in terms of processing load (as revealed by the highest times taken to accept it). This last result is particularly intriguing. According to the framework I am adopting, no general cost is to be associated to scalar implicature computation *per se*, contrary, for example, to Relevance Theory's predictions. A "cost" is instead to be expected if implicatures are added ("locally") in DE contexts, given that this would result in a loss of informativity. I believe that this finding need to be considered in the debate on the "cost" of SIs: if a "cost" were to be attributed to implicature computation *per se*, as Relevance Theory suggests, then, not only we should find for (b)-sentences in S2 (DC –only *inclusive*-condition for DE contexts) an analogous contrast in RTs to the one found for (a)-sentences in S1 (DC –only *exclusive*- condition for DE contexts), but we should also get significantly higher RTs in accepting (a) sentences in S1 than accepting (b) in S2. As we saw, none of these comparisons were significant. In this respect, these results seem to be in contrast with recent works on SI computation realized within Relevance Theoretic tradition (e.g. Noveck and Posada (2003), Bott and Noveck (2004), Breheny et al. (2005) and Katsos et al. (2005)). By means of different techniques, these authors conducted on-line experiments with adults evaluating sentences containing scalar terms in different settings. Very generally, their results seem to point to the same direction, namely: whenever subjects compute SIs, they do it at a "cost", that is reflected by a slowdown in correspondence of the scalar trigger when measuring reading times or by an increased time to process the whole sentence. These results were uniformly interpreted as evidence of the "cost" of SIs. Without entering the details of each study, I would like to make some general considerations about their findings. In the first place, the slowdown could simply reflect a general attitude of "pragmatic" responders, as also suggested by Noveck and Posada (2003). Secondly, the possibility that a strategy is involved is also attested by subject's distribution: in most (if not all) cases subjects split when they have to judge an underinformative sentence, especially when sentences are given "out of the blue", in the absence of a preceding context (a result also replicated here for condition (b)-S2), as if some participants consider the implicature "relevant enough" (to borrow from Relevance Theory terminology) and thus add it, while others don't. I believe that the solution proposed by Chierchia well explains these facts, being the activation of the alternatives optional, and also being the flip between having or not having the implicature in NON-DE contexts way easier than in DE contexts. On the contrary, it's more difficult to find a ready explanation of this split in subjects' distribution within Relevance Theory, given that the presumption of optimal relevance of a given stimuli should in principle be the same across participants in the same task.

To conclude, our findings seem compatible with Chierchia's "logicality" approach, which assumes that SI are computed "locally", as part of the recursive computational process and not via post grammatical operations, and that their derivation is regulated by a feature that (optionally) activates scalar alternatives but (mandatory) selects the most informative interpretation (which depends on monotonicity). Moreover, with respect to the theoretical issues explored above, we believe that our results may cast some doubts on the hypotheses that SI derivation is costly *per se*. Most importantly, they confirm the value of integrating theoretical claims in semantics with experimental work.

References

Bott, L., and Noveck, I. (2004) "Some utterances are underinformative: The onset and time course of scalar inferences", *Journal of Memory and Language* **51**, 433–456.

Breheny, R., Katsos, N., & Williams, J. (2005). "Are generalized scalar implicatures generated by defaults? An on-line investigation into the role of context in generating pragmatic inferences", *Cognition*, 1–30.

Chierchia, G. (2004) "Scalar Implicatures, Polarity Phenomena and the Syntax/Pragmatics Interface" in A. Belletti (Ed.), *Structures and Beyond*, Oxford University Press.

Chierchia, G. (2006) "Broaden Your Views: Implicatures of Domain Widening and the "Logicality" of Language", *Linguistic Inquiry* **37**(4), 535–590.

Foppolo, F. (2007) *The logic of pragmatics. An experimental investigation with children and adults*, Ph.D dissertation, University of Milano-Bicocca, Milan.

Fox, D. (2003). *Implicature Calculation, only, and lumping: another look at the puzzle of disjunction*, ms., MIT.

Grice, H. P. (1957) "Meaning", *Philosophical Review* **78**, 147–177 [in Grice 1989].

Horn, L. R. (1972) *On the Semantic Properties of Logical Operators in English.* Doctoral Dissertation, UCLA, CA.

Kadmon, N. and F. Landman (1993) "Any", *Linguistics and Philosophy* **16**, 353–422.

Krifka, M. (1995) "The Semantics and Pragmatics of Polarity Items", *Linguistic Analysis* **25**, 209–257.

Landman, F. (1998). "Plurals and maximalization", in S. Rothstein (ed.), *Events and grammar*, Dordrecht: Kluwer.

Levinson, S. C. (2000) *Presumptive Meanings – The theory of Generalized Conversational Implicatures*, Cambridge, MA: MIT Press.

Noveck, I., and Posada, A. (2003) "Characterising the time course of an implicature", *Brain and Language* **85**, 203–210.

Recanati, F. (2003) "Embedded Implicatures", *Philosophical Perspectives,* **17**(1), 299–332.

Russell, B. (2006) "Against Grammatical Computation of Scalar Implicatures", *Journal of Semantics* **23**(4), 361–382.

Sauerland, U. (2005) "On embedded implicatures", *Journal of cognitive science,* **5**.

Spector, B. (2003) *Scalar implicatures: local or global? Exhaustivity and Gricean reasoning*, paper presented at the Workshop on Polarity Scalar Phenomena and Implicatures, University of Milano-Bicocca.

Sperber, D., and Wilson, D. (1986) *Relevance: Communication and Cognition*, Oxford: Blackwell.