Decalog 2007: Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue, pages 125–131.
Trento, Italy, 30 May – 1 June 2007. Edited by Ron Artstein and Laure Vieu.

Between “cost” and “default”: a new approach to Scalar Implicature

Francesca Foppolo
University of Milano-Bicocca
francesca.foppolo@unimib.it

Abstract

Scalar Implicatures are pragmatic inferences that are normally derived in conversational exchanges when a scalar term, such as for example “or”, is used. Different theoretical accounts have been proposed to describe how and at which point in the derivation we actually add this inference. Large part of the most recent debate is focused on the question of the “cost” of implicature computation, an aspect that is crucial to choose among alternative accounts. In this perspective, my intent here is to present an experimental study in the ongoing debate centred on the “costly” or “default” nature of implicature computation. The main result of the study presented here is the fact that a “cost” is found only when the implicature is added despite the fact that it leads to a weakening of the overall assertion (namely, in DE contexts): this loss in informativity, and not implicature computation *per se*, is interpreted as the source of this “cost”. The theoretical background for this study is offered by Chierchia (2006) and his new intriguing parallelism between the phenomenon of scalar implicature and negative polarity.

1 The phenomenon

Scalar Implicatures (SIs henceforth) are pragmatic inferences that are normally derived in conversational exchange when a scalar term, such as “or” is used. Consider the example in (1) and (2):

(1) The dwarf is singing *or* dancing
(2) The dwarf is singing *and* dancing

What is normally conveyed by uttering (1) is that (2) doesn’t hold. This amounts to saying that, by uttering (1), the inference that the hearer is allowed to draw is (3), which is actually how a sentence like (1) is normally understood:

(3) The dwarf is singing *or* dancing <u>but not both</u>

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. In our example above, “or” belongs to an informational scale, i.e. <or, and>, in which “and” is the strongest element. By virtue of the fact that (2) constitutes the strongest alternative to (1) (it contains the stronger element “and”), and that (2) is not what was actually reported, then one is entitled to assume that (2) does not hold, hence the inference in (3) in which the negation of the strongest element on the scale is added.

2 The ongoing debate

Different theoretical accounts have been proposed to explain how and when implicatures are derived. We will focus here on one aspect of this debate in particular, namely the question of the “cost” of implicature computation. This has been the centre of the most recent debate between supporters of Relevance Theory (cf., a.o., Sperber and Wilson, 1986) on the one hand and of Default approaches on the other (cf., a.o., Levinson, 2000). The claim that implicatures are added at a cost by our processing system is necessary to differentiate these two approaches. In Levinsonian terms, implicature computation constitutes a default process, i.e. something that our computational/processing system performs automatically, thus it is by definition virtually costless. On the Relevance Theoretical view, instead, every operation imposed to our processing system must be evaluated in terms of “costs and benefits”, ultimately in terms of “relevance” to contextual assumptions: only those stimuli that are relevant enough are worth a processing effort. From this assumption, the claim that implicatures are costly necessarily follows: implicatures are only derived when explicitly required by the context, i.e. when the benefits that one gains from their computation overcome the processing effort required to derive them. If implicatures were costless, then the principle of optimal relevance would lose its foundation. This is the reason why all the experimental works on scalar implicatures within the Relevance Theoretic tradition have been focused on finding evidence of such a "cost".

Between these two approaches, there is a third proposal, recently delineated by Chierchia (2006). This approach seems to combine some features of the two approaches and, in my view, gives a new direction for solving the question of how and when and why scalar implicatures are derived. I will sketch this new proposal in the following section.

3 Chierchia's proposal

In Chierchia's most recent work (cf. Chierchia, 2006 in particular but also Chierchia, 2002), a unified account of negative polarity elements like *any* and scalar implicatures is being considered. In this new formulation, a binary feature σ is introduced as regulating the activation of scalar alternatives associated to scalar and negative polarity items. This feature can be assigned two values: [$\pm$ σ]. Selecting [$+\sigma$] results in the activation of the scalar alternatives (ALTs henceforth); selecting [$-\sigma$] results in the selection of the plain meaning in which ALTs are not active. The crucial point is that, whenever the feature [$+\sigma$] is selected, then the constraint on strengthening applies and an exhaustivization operator **O** (which has a meaning akin to that of *only*) must be used. For our purposes, it suffices saying that the result of this mandatory operation always leads us to the selection of the strongest – most informative – interpretation of the sentence containing the scalar item. With respect to the theoretical debate introduced in section 2, this new formulation leaves place to the notion of a *strategy* on the one hand and to the notion of *default* on the other: if the choice of activating the alternative interpretations of a statement containing a scalar term is in the end a matter of a subjective choice (thus, optional), once the selection has been made and the alternative interpretations activated, then the choice of the stronger alternative is instead mandatory. Very informally, the operator **O** applied to a sentence like (1) above, containing a scalar expression of the form "A or B" in which the ALTs are active will result in the derivation of the scalar implicature associated to *or*: **O** (singing $or_{[+\sigma]}$ dancing) = *only* (singing *or* dancing) = *only* (singing *or* dancing) and not (singing *and* dancing), thus excluding sentence (2) and deriving the inference in (3). The choice between activating the set of alternatives or not is considered optional in case of scalar terms while their activation is mandatory in case of NPIs. We won't pursue further the discussion on the parallelism with NPIs (this goes beyond the purposes of the present paper) but it's interesting to report a generalization on SIs already reported in Chierchia, 2002: "(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 (or Downward Monotone), i.e. those contexts that licence inferences from sets to their subsets. For example, the antecedent of conditional represents a canonical 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 tendency to be maximally informative and the monotonicity properties of the context, with respect to sentences (4), representing a DE context, and sentence (5), representing a NON-DE context, the distributional generalizations in (6) can thus be derived:

(4) If the troll is singing or dancing then he's happy (=DE)

(5) If the troll is happy, then he is singing or dancing (=NON-DE)

(6)

(a) The exhaustive interpretation (via application of the operator **O)** of a scalar term is easier in a NON-DE than in a DE context;
→ SI computation is easier in (5) than (4) (increased informativity)

(b) Having an implicature embedded in a DE context is way harder than having it embedded in a NON-DE context
→ SI computation is harder in (4) than (5) (loss of informativity)

(c) The flip between having an implicature and not having it is relatively easy in NON-DE contexts
(activation or de-activation of ALTs)

(d) The flip between having an implicature and not having it is hard in a DE context (loss of informativity)

These predictions have been specifically tested in the experimental study that I'm going to present in the next section.

4 A reaction-time study

As we have seen, Chierchia's proposal makes clear-cut predictions as to when the derivation of SIs is expected, also in relation to the type of syntactic context in which the scalar term operates. In this respect, and in consideration of the debate on the "cost" of SI computation reported in section (2), the experiment I'm going to present addresses the following questions:

(7)
(i) whether one of the candidate interpretations constitutes the preferred one depending on the syntactic environment it appears in (to this purpose, the rate of acceptance/rejection of critical sentences across conditions will be considered);
(ii) whether the derivation of the implicature is a costly process (to this purpose, the analysis of reading times (RTs henceforth) will be analysed).

4.1 Participants

A total of 30 subjects participated in this experiment. Participants were mainly 1st year students at the Psychological Faculty of the University of Milano-Bicocca, and received credits for their participation.

4.2 Procedure

The experiment was realised using E-Prime. Subjects were tested in a quiet room using a laptop and after a training session. Participants' task was to evaluate sentences as "true" or "false" with respect to a scenario constituted by four pictures that appeared on the screen. After an introductory screen in which characters and objects were presented for the first time, critical material was presented as follows (by pressing the space bar on the keyboard): at the top of a black screen a sentence appeared (in white). Participants were instructed to read (silently) the sentence and then press the space bar key to see the four pictures describing the situation against which they had to evaluate the sentence as "true" or "false". By pressing the bar, the four pictures appeared on the screen in the space below the sentence (in a random order). To answer, subjects had to press one of two highlighted keyboard keys: a green key for "true" and a red key for "false". After pressing it, they were either asked to move on by pressing the space bar (whenever their answer was "false") or, in case they answered "true", they had to answer another question that appeared in the middle of the screen (the four pictures remained there): "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). Only time to answer the True/False question was recorded, starting from the moment they pressed the bar to make the pictures appear on the screen till they pressed the True/False key. Critical conditions were treated as a within subject factor: 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 fillers.

4.3 Material

To avoid interferences from extra-linguistic factors on the interpretation of sentences, all the material presented in this experiment contained only fantasy names for characters and objects. Characters were in fact introduced as inhabitants of weird planets with their bizarre objects, unfamiliar to inhabitants on Earth.

The experiment presented a 2×2 condition design, in which two conditions were created as a within subject factor, each displaying 2 different levels. Condition I represents the type of syntactic environment in which the disjunction appears. The monotonicity properties of the context is varied, as summarized in (8): in sentences of type (a) *or* is embedded in the antecedent of conditional, which crucially constitutes a DE environment, like (4); on the contrary, in sentences of type (b) *or* is embedded in the consequent of the conditional, which constitutes a NON-DE environment like (5) above.

(8) Condition I: monotonicity of the context
(a) *If a P has an <u>A or a B</u>, then he also has a C* [= DE context]

(b) *If a P has a C, then he also has an A or a B* [= NON-DE context]

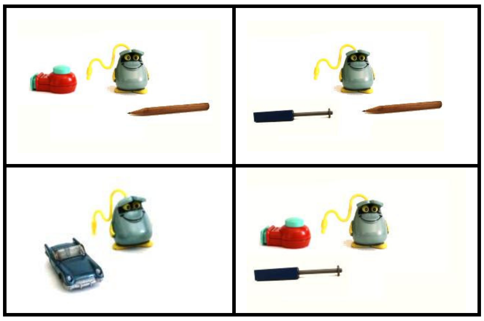
Each sentence was presented to each subject in two different critical situations, corresponding to levels S1 and S2 of Condition II (see (9) below). Each situation modulated the interpretation associated to the scalar term contained in the sentences by means of the scenario represented by the set of four pictures.

(9) Condition II: situations
S1: a situation representing the *exclusive* interpretation of *or*;
S2: a situation representing the *inclusive* interpretation of *or*.

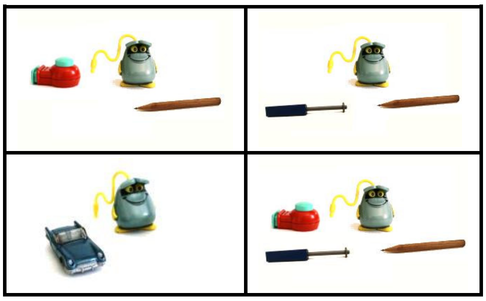
Consider, for example, the following test sentences (recall that fantasy names were used):

(10)
(a) If a Glimp has a curp or a dorf, then he also has a pencil
(b) If a Glimp has a pencil, then he also has a curp or a dorf

These were tested (on different subjects) in the following scenarios, representing conditions S1 and S2 respectively:



S1: only compatible with *exclusive* interpretation of *or* (see last picture: A and B but not C).



S2: only compatible with *inclusive* interpretation of *or* (see last picture: A and B and C)

Please note that the only crucial difference between the two scenarios is represented by the last picture in the sequence (remember that, during the experiment, the order of presentation of the four pictures was completely randomized across items and subjects). Crucially, scenario S1 is only compatible with the *exclusive* interpretation of *or*, which is the most informative in case of sentences of type (b), i.e. in a NON-DE context, but not of sentences of type (a), i.e. in a DE context. On the contrary, scenario S2 is only compatible with the *inclusive* interpretation of *or*, which is the most informative in case of sentences of type (a) but not of sentences of type (b).

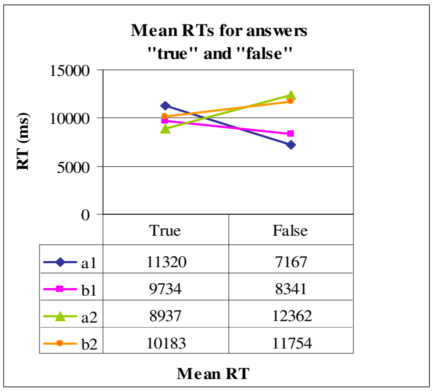
4.4 Results

Results are summarized in the table below, divided per type of sentences which crucially differ in their monotonicity properties: the 2nd column reports the type of scenario in which the sentence is evaluated (recall that S1 corresponds to the *exclusive* interpretation of *or* while S2 corresponds to the *inclusive*); the 3rd column reports the percentage of "true" answers followed by the rate assigned to the scale that appears in the 4th column; the last three columns report respectively: the response times to answer "true", to answer "false" and the mean total response time per condition.

Sent.	**Sit.**	**True**	**Scale rate**	**RTs for True**	**RTs for False**	**Mean RTs**
(a) DE	S1 (exc)	**57%**	3.47	11320	7167	9628
	S2 (inc)	**90%**	3.81	8937	12362	9291
(b) NON -DE	S1 (exc)	**87%**	4.38	9734	8341	9549
	S2 (inc)	**77%**	4.04	10183	11754	10562

Data on critical items can be analyzed with respect to different parameters: percentage of "true" and "false" answers; time taken to make a decision between "true" and "false"; grade assigned to the scale. I will focus here on the main findings. First of all, a large majority of subjects (90%) accept (a) sentence in Condition S2, compatible with the *inclusive* interpretation of *or*, while only half of them (57%) accept it in S1, where *exclusive* interpretation of *or* is represented. This difference is statistically significant (t(29)=-3.34, p<.01). In the second place, the rate of acceptance of the (a)-sentence in Condition S1 (representing the *exclusive* interpretation) is also significantly different from the rate of acceptance of the (b)-sentence (representing a NON-DE context) in the same condition (57% vs. 87%, t(29)=-3.07, p<.01). Moreover, those subjects that accepted the sentences in scenario S1 assigned a significantly lower score to (a) than (b) sentences (t(41)=-2,59, p<.01).

Data reported in the Table above are also interesting in another respect: reaction times to evaluate critical items in different conditions can be compared, considering overall mean RTs per sentence-type or distinguishing between RT to answer "true" and "false" separately, as plotted in the graph below.



A first point worthy of remark is the fact that no significant difference emerges taking context (DE vs. NON-DE) or scenario (*inclusive* vs. *exclusive*) as critical factors. These results seem to indicate that the processing load required to evaluate both types of sentences in both conditions was almost identical, at least if we consider mean RT overall. However, this consideration should be handled with care, given that one needs to integrate the overall picture with the data plotted in the graph, showing RTs for both sentence types and situations but differentiated between "true" and "false" type of answer.

Let's discuss Relevance Theory predictions first. According to this approach no difference due to the monotonicity properties of the two contexts is in principle to be expected. In fact, according to that approach, analysis of RTs should reveal a "cost" of scalar implicature computation. In this respect, the first crucial comparison is the one between RTs for answering "true" between situations S1 and S2 and a comparison on RTs for answering "false" between the same conditions. The second comparison to reveal the "cost" of implicature is the one between the RTs for answering "true" and the RTs for answering "false" within the same condition. None of these comparisons, however, turned out statistically significant.

Most interestingly, among RTs, only one comparison revealed statistically significant. Precisely, this was the time to answer "true" in situation S1 in case of sentence (a) compared to the mean time to answer "false" when evaluating the same sentence in the same condition (t(29)=5.16, p<.001). This reflects the fact that subjects that derived the implicature in case of DE context did it at a "cost". This finding is crucial in two respects: the same presumptive "cost" did not emerge from any other comparison, contrary to the Relevance Theory's expectations; also, this was the only "hard step" predicted by the distributional generalizations outlined in (6) derived from Chierchia's theory.

4.5 Discussion

One of the questions addressed in this experiment was the influence of the syntactic context on SI computation, ultimately the effect of the monotonicity properties of the context on informativity. Considering the acceptance rate first, we can say that the results obtained confirm our predictions. In the first place, subjects treat the two sentence types differently in the two situations; secondly, they derive SIs more when *or* appears in a NON-DE than in a DE context; thirdly, they prefer not to derive the SI when *or* appears in a DE context. The second question raised in (7) above asked whether the process of computing implicatures is costly. According to the framework I am adopting, no cost is to be associated to scalar implicature computation *per se*, contrary to Relevance Theoretic approaches. A cost is instead to be expected when the implicature is derived despite the fact that the scalar term is embedded in a DE context, in which the adding of the implicature would result in a weakening of the overall assertion. This prediction seems largely supported by the results: only those participants that accept the (a)-sentence in S1, thus deriving the implicature in a DE context, took significantly longer than the participants that reject that sentence in the same condition. If the cost were to be attributed to implicature computation in general, then the same contrast should be found in case of sentence (b), but this is not so. To account for the data obtained in this task, the claim that implicature computation *per se* is costly is, in my opinion, to be rejected.

In summary, the claim that the default interpretation of the scalar term depends on the monotonicity properties of the context in which the scalar term is embedded is largely supported by the data obtained in this experiment: without such a claim, it would be difficult to account for the fact that sentence (a) in which *or* appears in a DE context, when evaluated in scenario S1 representing the *exclusive* interpretation, is the hardest condition of all, both in terms of subjects' distribution, scale rate and RTs.

5 Conclusions

The experimental results presented here seem to be in contrast with recent works on SI computation realized within the Relevance Theoretic tradition. In particular, I'm referring to the works by 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 on adults while evaluating underinformative sentences containing scalar terms such as *some* and *or* in different experimental "situations" (as for, e.g., presence or absence of a preceding biasing context, or different instructions/suggestions given to participants to fulfil the task). Very generally, the results of these studies seem to point to the same direction, namely: whenever subjects compute the implicature associated to a scalar term, they do it at a cost. This is reflected by a slow down in correspondence of the scalar trigger when measuring reading times (like in the studies presented by Breheny and colleagues and Katsos and colleagues), or by an increased time to process the whole sentence (when measuring reaction times, like in the Bott and Noveck's study or in the ERP study conducted by Noveck and Posada). These results were uniformly interpreted by these authors as evidence of the fact that scalar implicature computation is a costly process. Without entering too much in the details of each single study, I would like to make some general considerations about the findings of the works mentioned above. In the first place, let's consider subject's distribution. It's interesting to note that in most (if not all) cases subjects split when they have to judge an underinformative sentence, even when the sentence is given "out of the blue", i.e. in the absence of a preceding context (this finding was also replicated in the experiment presented here). This is a clear indication, according to my view, that subjects are adopting a strategy to which they stick when solving the experimental task: half of the subjects consider the computation of the implicature "relevant enough" (to borrow from Relevance Theory terminology) and thus add the implicature; the other half, instead, keep to the plain meaning of the scalar term, and do not derive the implicature. I believe that the solution proposed by Chierchia (2006) well explains these facts, being feature selection the result of a subjective choice, 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 the Relevance Theory given that the presumption of optimal relevance of a given stimuli should in principle be the same across all participants.

On the other hand, RT data seems at first glance to be better explained by Relevance Theory. The crucial comparison, according to this approach, is between the RTs of the two groups: subjects that derive the implicature always take longer than the rest. This result is sufficient, according to them, to claim that the process of computing SI is costly and thus subjects only derive SIs when the benefits obtained by the adding of a SI exceed the processing effort required for its derivation. However, it's not that clear that this overload is effectively due to the adding of the implicature *per se*. As the results in the experiment presented here show, the only "cost" is found when the implicature is added despite the fact that it leads to a weakening of the overall assertion. As we said, this "flip" is predicted to be hard in Chierchia's generalization and this loss in informativity, and not implicature computation, seems to be the source of this "cost".

In the end, I believe that the intriguing debate on pragmatic inference, which has very recently attracted the interest of psycholinguists, is far from being solved. To begin with, the majority of the studies have been focused on measuring the "cost" of implicature derivation. Though interesting, I think this is not the *only* question to be solved within a semantic-pragmatic theory of Scalar Implicatures.

References

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

Breheny, R., N. Katsos, et al. (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. (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.* Unpublished Ph.D. dissertation, University of Milano-Bicocca.

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

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

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