Interaction of Structural and Contextual Constraints During the On-line Generation of Scalar Inferences

Napoleon Katsos (nk248@cam.ac.uk)
Research Centre for English and Applied Linguistics,
University of Cambridge, 9 West Road, CB3 9DP, UK

Richard Breheny (richardb@linguistics.ucl.ac.uk)
Department of Phonetics and Linguistics, University College London
Gower Street, WC1E 6BT UK

John Williams (jnw12@cam.ac.uk)
Research Centre for English and Applied Linguistics,
University of Cambridge, 9 West Road, CB3 9DP, UK

Abstract

We investigate the interaction of structural and contextual constraints on the on-line generation of three types of Scalar Implicature, (a) the disjunction: "A or B" >> "either A or B but not both", (b) the partitive NP: "some of the Fs">> "at least one but not all of the Fs" and (c) ad hoc context dependent scales. In theoretical linguistics, according to the structural approach (Chierchia, 2004; Levinson 2000) Scalar Implicatures (SIs) are generated whenever certain constraints on the semantic properties of the linguistic structure are satisfied. According to the pragmatic approach, structural properties are a necessary but not sufficient condition, since for SIs to be generated there ought to be further contextual constraints (Carston, 1998; Recanati, 2003; Sperber & Wilson, 1995). The linguistic debate can be addressed experimentally, but studies in this small but growing area are not reaching a consensus. Single sentence truth value judgment tasks (Bott & Noveck, 2004; Noveck & Posada, 2003) and studies on the visual-world eye-tracking paradigm (Storto & Tanenhaus, 2005; Huang & Snedeker, 2005) support the pragmatic approach but do not explicitly manipulate discourse context constraints. Text comprehension studies that do so (Bezuidenhout & Cutting, 2002; Bezuidenhout & Morris, 2004, Breheny, Katsos & Williams *in press*) have been ambivalent. We present one off-line and three on-line text comprehension experiments that address the previous shortcomings and conclude in favor of the pragmatic approach. The linguistic debate is related to the psycholinguistic discussion of the factors that constrain sentence processing. We discuss our findings with regards to the modular versus interactive nature of the human parser.

Keywords: pragmatics; scalar implicature; processing.

Introduction

Certain linguistic expressions form entailment scales where terms on the right of the scale are informationally stronger than terms on the left (measured by number of entailments, e.g. <some, most, all>, <or, and>). Scalar expressions may trigger Scalar Implicatures (SIs) in upward entailing structures when the assertion of an informationally weaker term conversationally implies (">>") the negation of the stronger terms in (1).

(1)   Mary: How many students failed the test?
        Jane: Some of them did >> Not all of them did

A distinctive property of SIs is defeasibility due to structural and contextual constraints. Regarding the former, SIs are dependent on the direction of the entailment of the scale. In downward entailment structures, e.g. in the antecedent of a conditional, the direction of entailment in the scale is reversed ("some" entails "all") and SIs are not generated:

(2)   If some of the students fail the test, their teacher will be disappointed.

Here there is no implication that if all the students fail the test their teacher will not be disappointed.

Moreover, SIs are context dependent. There are three types of relevant contexts, (a) upper bound, where it is necessary to generate an SI in order to satisfy a discourse goal like in (1), (b) neutral contexts, where there is no assumption as to whether the SI is relevant or not, and (c) lower bound contexts, where the discourse goal can be satisfied by the plain meaning of the SI trigger without any need for the SI. It is assumed that SIs are not generated in lower bound contexts since the assertion of stronger terms of the scale is irrelevant:

(3)   Mary: Why is the teacher disappointed?
        Jane: Some of his students failed the test.

In (3) the discourse goal that is set by Mary's question is satisfactorily addressed by the plain meaning of "some", "at least one of the students and even all", so it becomes irrelevant whether the strengthened alternative "at least one but not all" is true or not (e.g. Levinson, 2000).

The linguistic debate

The current linguistic debate is set between the structural approach (Chierchia, 2004; Levinson, 2000) which posits that logical properties of the linguistic structure are the necessary and sufficient conditions for SI generation, and the pragmatic approach (Carston, 1998; Recanati, 2003; Sperber & Wilson, 1995) that posits an interaction between the three types of

context (upper-, lower-bound and neutral) and structural constraints.

The structural approach

According to Chierchia's version, SIs are generated locally, below the level of the full sentence, and automatically, as soon as a scalar term enters the computation. The grammar assigns two values for the scalar trigger, the plain meaning of the expression and its scalar alternative which incorporates the SI. Subsequently, the informationally weaker meaning will be filtered out. In the case of upward entailing structures this is the plain meaning of the trigger, but in downward entailing structures this is the meaning with the SI. Thus, there are cases where SIs are computed by default and cases where they are absent by default; in both cases the key factor is structural properties. At a second stage, the SI is checked against the background assumptions. If the addition of a default SI to a given discourse context is infelicitous, (e.g. to a lower bound context as in 3), the SI will be cancelled; but this happens only at a later stage after compositional grammatical processes have been concluded. A corollary of the structural approach is that SIs are linguistic inferences generated by a default process that should come for little processing cost.

The pragmatic approach

According to the pragmatic approach the structural constraint of upward entailment is indeed a necessary condition for SI generation and SIs will not be generated in downward entailing structures. However, for SIs to be generated, besides being in an upward entailing structure, they must also be under an upper-bound discourse context, i.e. a discourse goal which makes the SI relevant (e.g. for the SI trigger "some of the Fs" this would be a discourse where the quantity of the Fs involved is important).

The differences between the structural and the pragmatic approach become evident in upward entailing structures (the structures that allow for SI generation) in two respects: First, given lower-bound discourse contexts where the SI is incompatible with the discourse, the first predicts that SIs will be generated and subsequently cancelled and the second that SI will not be generated at all. Moreover, in upward entailing structures which are discourse neutral (i.e. there are no contextual assumptions that make the SI relevant), the structural approach predicts the generation of SIs regardless of the absence of contextual assumptions whereas the pragmatic approach predicts that the SI will not be generated. Table 1 explicates the predictions.

Psycholinguistic investigation

Chierchia (2004), Levinson (2000), Sperber & Wilson (1995) and other linguists have explicated the need for relevant psycholinguistic evidence. There is a small but growing literature employing various methodologies.

Text comprehension

The study reported in Bezuidenhout & Cutting (2002) investigates prediction (v), the case where an SI in an upward entailing structure is incompatible with a

Table (1). Structural & Pragmatic Approaches' Predictions

	Structural Approach	Pragmatic Approach
(i) SIs generated locally	YES	YES
(ii) SIs are not generated in downward entailment structures	YES	YES
(iii) In upward entailing structures with an upper-bound discourse context	SI generated	SI generated
(iv) In upward entailing structures which are discourse neutral	SI generated	SI NOT generated
(v) In upward entailing structures with a lower-bound discourse context	SI generated & cancelled	SI NOT generated
(vi) In terms of processing resources	generation costless & cancellation costly	generation costly & cancellation never occurs

lower-bound discourse context. Bezuidenhout & Cutting (hence B&C) test various scales including the numerals and quantifiers. According to the structural approach, in a lower-bound context the SI has to be cancelled, leading to longer (or at least equal) reading time than in an upper-bound context where the SI is allowed to be generated. In a self-paced reading experiment B&C found that there was a significant slowdown when reading the SI triggers in lower-bound contexts compared to upper-bound. They conclude that in lower-bound contexts SIs are first generated and then cancelled, supporting the structural approach.

However, we must note that B&C's experimental items were not "orthodox" implicatures according to the current consensus in linguistic theory. The cardinals by which they exemplify their case are not considered implicatures by most theorists in either camp of the debate (see Breheny, 2005; Geurts, 1998 i.a.; Papafragou & Musolino 2003 for developmental evidence). Other expressions B&C used have never been considered implicatures. E.g. inferring from 'everyone is a vegetarian' that 'everyone in this group (= not in the whole universe) is a vegetarian' is a process of narrowing down the range of application of the quantifier and not of implicature generation. All in all, almost half of the categories B&C tested do not give rise to SIs.

Bezuidenhout & Morris (2004; hence B&M) used an eye-tracker to record reading time and gaze fixations on sentences like (4a) where the 'not all of the Fs' SI is explicitly cancelled in the second sentence by the phrase 'in fact all of them did'. Definite expressions were used as controls (4b):

(4)   a. Some books had colour pictures. In fact all of them did, which is why the teachers liked them.
         b. The books had colour pictures. In fact all of them did, which is why the teachers liked them.

According to the structural approach, participants should generate the SI when 'some' is read and will engage in

cancellation and reanalysis at the 'them did' region of the sentence. In contrast, according to B&M, the pragmatic approach claims that participants do not fully commit to the 'some but not all' SI right away. This approach predicts increased processing time on the word 'all' to reflect the fact that the reader has registered information potentially relevant to the specification of an underspecified item.

B&M find that there is no reading time slowdown at the 'them did' region between (4a) and (4b) but there is a slowdown on 'all' in (4a), which they interpret as evidence in favor of the pragmatic approach.

There are however a number of objections. Firstly, the discourses used were not unbiased since a) at the beginning of the second phrase there always was a discourse marker which indicated contrast ('in fact', 'actually' etc) which is not neutral to the inferences under study and b) their control example (4b) is an infelicitous discourse and further questions arise from a follow up study that was designed to control for this. Second, B&M's understanding of underspecification is problematic, implying that interpretations can remain underspecified indefinitely in case there is no further relevant information in the discourse. It is our understanding that both structural and pragmatic approaches clearly ascribe to the prediction that if an SI is generated then this should occur immediately once the trigger has been processed. All in all, the studies by B&C and B&M have marked the beginning of SI experimental investigation but for different reasons both studies remain inconclusive.

Single sentence truth value judgment tasks

When presented with an SI trigger and asked to make a judgment, respondents may answer "false" to a sentence like (5) if they generate the SI ("at least one but not all") or answer "true" if they use the plain meaning ("at least one and perhaps all"):

(5)   Some elephants have trunks

In a reasoning experiment, Noveck & Posada (2003) instructed groups of participants to use scalar terms either with the SI or just with the plain meaning. They timed the responses in a sentence verification task and found that the time taken by the participants answering 'false' was significantly longer than those answering 'true'. That is, when respondents based their answers on the plain meaning, it was not the case that the SI was first generated and then cancelled but rather that the SI was not generated in the first place. This result supports a model where SIs are costly inferences. In a similar study Bott & Noveck (2004) replicated these findings and also excluded the possibility that the difference was due to the difference in response type (True vs. False) by including appropriate control conditions. Moreover, in a second experiment they introduced an additional layer of narration, where the sentences are attributed to a character. The stimuli are preceded either by the declaration that "Mary says the following sentence is true" or that "Mary says the following sentence is false". Participants are then asked to agree or disagree. Bott & Noveck replicated the previous findings with a design where it is possible for answers based either on the SI or the plain meaning to have the same

response type. In the third experiment they removed the instructions to the participants group and still obtained the same pattern and in the fourth experiment they found that the number of responses based on the SI increased as permitted response time increased. Thus, the studies by Noveck and colleagues have yielded consistent evidence against the structural approach regarding prediction (*v*) and prediction (*vi*).

Visual-world eye-tracking

Storto & Tanenhaus (2005) investigated the scale <or, and>, whereby the plain meaning of the disjunction "or", called the inclusive interpretation, is "A or B or even both" and the SI gives rise to the exclusive interpretation "either A or B but not both". Storto & Tanenhaus found that the SI was calculated and integrated very locally to the utterance of the disjunction and could guide the further processing of the sentence. Thus they supported prediction (*i*), which is shared by both the structural and the pragmatic approach, that the process that generates SIs operates quickly, below the level of the full sentence. However, the facilitation from the generation of the SI was not as local as the facilitation obtained from the meaning of the stronger term of the relevant scale, 'and'. It is possible that the delay in facilitation from an SI relative to facilitation from semantic meaning may be due to some shortcoming of the design – and the authors are cautious not to exclude this possibility. However, taken at face value, this result suggests that SIs, however local, are not on a par with semantic meaning.

Huang & Snedeker (2005) compared the processing of the partitive noun phrase "some of the F" and of number terms using the visual-world paradigm. Their results suggest that numerals (whose interpretation is grammar-driven) are categorically different from scalar terms proper. When presented with a set of plausible alternatives in the number condition (one, two or three Fs) participants fixated on the correct image from the beginning and engaged in very early disambiguation of the sentence. In contrast, in the scalar condition, participants were presented with two alternatives ("some of the Fs" and "all the Fs"), initially looked at both plausible alternatives, and showed relatively late disambiguation. These findings suggest again that even though SIs are generated on-line locally, the time-course of the process that generates them is different from that of the process that generates grammar driven interpretations like the numerals.

Overall, single sentence truth value judgment tasks and visual-world eye-tracking studies suggest that SIs are local but not default inferences. However, these studies do not manipulate discourse context (upper-, lower-bound and neutral) which would allow for testing the crucial prediction (*v*) where the pragmatic and the structural approach make differential claims.

The present research

We aimed to devise experiments that would put to test all six predictions of the structural and the pragmatic approach, by embedding SIs in discourse contexts and manipulating the interaction with structural properties.

Experiment 1. An off-line study on defeasibility. In our first experiment, we tested the assumption of both the structural and the pragmatic approach that SIs do not arise (a) in upward entailing structures with lower-bound contexts and (b) in downward entailing structures. Even though this is a critical assumption, according to our knowledge it has not been experimentally addressed up to now. We tested the SI associated with the exclusive interpretation of the disjunction "A or B" >> "either A or B but not both". We created pairs of sentences where the first sentence was either a question which requires a single entity for an answer, therefore creating an upper-bound context, or a question which can be answered by a list of entities, creating a lower-bound context. In the upper- and lower-bound condition the second sentence was identical and contained a disjunction in an upward entailing structure that answered the preceding question. In the downward entailing condition the question in the first sentence was upper-bound, but the second sentence embedded the SI trigger in the antecedent of a conditional (which is a downward entailing structure). We ran an off-line sentence completion task where the final verb of the second sentence was missing. Participants were asked to fill it in using a closed class of verbs (is/are, does/do, has/have) all of which are marked for number in English. Our assumption is that if participants generate the SI that "not both A or B did so", they will use a verb to agree with the noun phrase "A or B" in singular number. If they do not generate the SI but use the plain meaning of the disjunction "A or B or even both", they will be much more prone to use a verb that will agree with the disjunction in plural number.

Design and results. We created 18 critical items each in 3 conditions (upper-bound, (UB), lower-bound, (LB), and downward entailing, (DE), (6)) which were rotated in 3 presentation lists (total n of participants = 45, adult speakers of British English). 18 fillers were added which had a similar structure to the critical items, but at the place of the SI trigger they contained only one name (e.g. "Jones").

(6)   **UB:** The manager asked: Who has the report on last year's profits? Her secretary replied: Jones or Barnes from the department of Finance…
        LB: The manager asked: Who has a report on last year's profits to show me? Her secretary replied: Jones or Barnes from the department of Finance…
        DE: The manager asked: Who has the report on last year's profits? Her secretary replied: If Jones or Barnes … it, I will bring the report on your desk in a few minutes.

Participants used a verb in singular agreement 82% in upper-bound, 52% in lower-bound and 51% in downward entailing. There was a main effect of Condition (F (2, 39)= 52.75, $p <$ 0.001) indicating that participants built an exclusive interpretation in the first case and an inclusive in the latter two. It is suggested that indeed SIs are generated in upward entailing contexts but not generated in downward entailing structures, as both the structural and the pragmatic approach assume in predictions (*ii*) and (*iii*). SIs are not generated in lower-bound contexts either (*iv*). However, from this study

we cannot know whether the end result (no SI) is achieved through the generation and subsequent cancellation of an SI or simply without the generation of the SI in the first place. This issue is raised in prediction (*v*).

Experiment 2. On-line processing in upper- and lower-bound contexts. To test prediction (*v*) we ran an on-line reading time experiment. The structural approach predicts that the SI trigger will be read *faster* (or at least as quickly) in the upper-bound than the lower bound context (where the SI has to be generated and then cancelled) and the pragmatic approach predicts that it will be read *slower* in the upper-bound context (because in the lower-bound the SI is not generated at all).

Design and results. We used the same18 critical items that we employed in Experiment 1 in two conditions, the upper- and lower-bound context. We completed the missing verb in the second sentence of each item with the appropriate verb in the singular, which is an acceptable form in either condition (over 50% in Experiment 1), as in (7). The items were rotated in 2 presentation lists (total n of participants = 30 adults speakers of British English). 60 items which were the critical items of unrelated experiments were added as fillers. Reading times were recorded in a segment by segment self-paced reading paradigm whereby the phrase that contained the disjunction appeared as one segment on the screen. Comprehension was motivated by questions.

(7)   **UB:** The manager asked:/ Who has/ the report/ on last year's profits?/ Her secretary replied: /Jones or Barnes/ from the department of Finance/ has. Would you like/ to see the report?
        LB: The manager asked:/ Who has/ a report/ on last year's profits/ to show me?/ Her secretary replied: /Jones or Barnes/ from the department of Finance/ has./ Would you like/ to see the report?

The critical segment, "Jones or Barnes", was read in 819ms in the upper-bound and in 775ms in the lower bound condition. There was a main effect of Condition, F (1,38)= 5.05, $p <$ 0.03. No other significant differences were recorded. In line with the results of Noveck and colleagues, it is suggested that it wasn't the case that in the upper-bound condition the SI was first generated and then cancelled, but rather that it wasn't generated in the first place. Moreover, SIs are generated locally, before the end of the sentence (as both approaches predict, *i*) in agreement with the findings of Storto & Tanenhaus (2005) and Huang & Snedeker (2005). It is also suggested that computing an SI is actually costly, as reflected in reading time required (prediction *vi*).

Experiment 3. On-line processing in ad hoc scales. In this experiment we replicated the finding that SI generation is a local and a time-consuming process by using non-logical scales introduced by the discourse context (e.g. <roof, house>, <father, parents>). We created 12 items in upper-and lower-bound conditions (8) which were rotated in two presentation lists (n of participants = 30). Another 50

unrelated items were added as fillers. Comprehension was motivated by questions.

(8) **UB:** George went/ to pick up Mary/ from the station./ He was covered/ in paint./ Mary asked him:/ Were you painting/ the house?/ George replied:/ I was painting/ the roof/ with an insulating paint.
LB: George went/ to pick up Mary/ from the station./ He was covered/ in paint./ Mary asked him:/ What were you painting?/ George replied:/ I was painting/ the roof/ with an insulating paint.

In a self-paced reading time paradigm the UB condition was read in 677ms and the LB in 623ms. There was a main effect of Condition (F (1, 16)=8.24, $p < 0.02$) for the critical segment (and a marginal spillover effect ($p < 0.09$) in the next segment), in line with the findings of Experiment 2.

Experiment 4. On-line processing in neutral contexts. In a previous study, Breheny, Katsos & Williams (in press) created discourse neutral contexts by presenting pairs of sentences without preceding context. We manipulated a discourse constraint, the sentence's information structure, and put the SI trigger "some of the Fs" in either Topic or Non-Topic position in the first sentence. We thus introduced implicit assumptions about how relevant the group of Fs is. The second sentences contained the anaphoric expression "the rest of the Fs" whose interpretation is facilitated if participants have inferred that "some of the Fs G-ed" implies that "not all the Fs G-ed". We compared reading time in each condition with a control condition where the SI was made explicit by the operator "only", which introduces alternative sets on-line (Ni, Crain, & Shankweiler 1996).

Design and Results. 24 critical items were created in 4 conditions (Topic, Topic & Only, Non-Topic, Non-Topic & Only, see 9) and rotated in 4 presentations lists. The experiment was run in Greek, which has a flexible word order and the initial position on the sentence is consistently associated with topicality. 48 participants took part. Comprehension was motivated by questions.

(9) **T:** Some of the consultants/ met the manager./ The rest/ did not manage/ to attend.
T-Only: Only some of the consultants/ met the manager./ The rest/ did not manage/ to attend.
Non-T: The manager met/ some of the consultants./ The rest/ did not manage/ to attend.
Non-T-Only: The manager met/ only some of the consultants./ The rest/ did not manage to attend.

Table (2). Mean reading time and SD of Experiment 4

	Reading time (ms)	SD
Topic	613	125
Topic & Only	611	110
Non-Topic	628	138
Non-Topic &Only	586	112

The critical interaction between position (Topic/Non-Topic) and the presence of "only" was significant, by subjects $F_s(1,44) = 4.24$, $p < 0.05$; and by items $F_i(1,19) = 6.93$, $p < 0.05$. Reading times were slower in the Non-Topic than the Non-Topic & Only condition, but there was no such difference between the Topic and the Topic & Only conditions. It is suggested that contrary to the structural approach (prediction *iv*), the SI was only generated when the trigger phrase was in one of the topic conditions, even though both the topic and the non-topic positions were in structurally identical sentences (upward entailing).

In this paper, we investigated whether the slow reading times in the Non-Topic condition obtained by Breheny, et al. (in press) could be due to a potential topic shift between the trigger and the target sentence. When participants read the second sentence they might assume that the topic of the first sentence, the first NP mentioned, will be the preferred continuation. The target phrase "the rest" refers to the second NP, thus violating topic continuity. If this potential topic shift is avoided in the case of the Non-Topic & Only due to the presence of "only", then the slow down in the Non-Topic condition may not be reflecting SI generation but a topic shift penalty. We ran an off-line sentence continuation task with the items used in Experiment 4 with 20 participants. In the Non-Topic condition the continuations referred to NP1: 30%, to NP2: 52,5%, and OTHER: 17,5%. In the Non-Topic & Only to NP1: 24,2%, to NP2: 59,2%, and OTHER: 16,6%. Crucially, both conditions prefer an NP2 continuation over an NP1 (Non-Topic: F (1, 16)= 5.01, $p < 0.04$; Non-Topic & Only: F (1, 16)= 12.09, p < 0.003); their difference in this preference is not significant (F (1, 16)= 0.022 n.s.). Therefore neither condition is paying a topic shift penalty. However, we also ran a second on-line follow-up study similar to the original except that the target phrase "the rest" was in non-topic position in the second sentence:

(10) **Non-topic (Only):** The manager met/ (only) some of the consultants./ He will talk/ to the rest/ tomorrow.

The results replicated the original study that SIs are not generated in the Non-Topic condition even though this condition is structurally identical to the Topic one.

Discussion

In this paper we presented one off-line and three on-line text comprehension experiments that tested 6 predictions concerning the processing of Scalar Implicatures. Experiment 1 provides evidence in support of the fundamental assumptions shared by both structural and local approaches, namely that SIs can be defeated by discourse context and structural properties. Experiments 2, 3 and 4 provide evidence in favor of the pragmatic approach, which predicts an interaction between structural properties of the sentence and discourse context. The conclusion that SIs are pragmatic, context-dependent inferences is aligned with the work of Noveck and colleagues on single sentence truth value judgment task and work on the visual-world eye-tracking paradigm (Storto & Tanenhaus, 2005; Huang & Snedeker, 2005).

The debate we address is one which will be familiar from sentence processing concerning whether there is an initial stage of processing where an encapsulated type of information (usually structure-based) operates first, and other types of information are considered later on, potentially causing back-tracking and re-analysis, a phenomenon known as Garden Pathing (Ferreira & Clifton, 1986 i.a.). The alternative view is that different types of information from syntax, semantics, discourse context, the situation referred to and statistical frequency of co-occurrence among other, interact from the earliest possible stage (MacDonald et al 1994; Spivey-Knowlton & Sedivy, 1995 i.a.).

The structural linguistic approach surveyed here is compatible with a modular and serial parser, whereby SIs are generated automatically by the encapsulated grammar – and only subsequently affected by information from other modules. As we have seen, our results, as well as those of Noveck and colleagues, are incompatible with such a model, since no effect of reanalysis was detected.

However, our findings are consistent with two alternatives. First, it is conceivable that language processing is modular, however SIs are not derived in the grammatical module, but in the pragmatics module which would operate on the output of the grammar. This would explain why there is no cancellation of SIs in lower-bound contexts, and why SIs are time-consuming inferences when compared to semantic inferences. On the other hand, our findings do not exclude the possibility of an interactive process, where statistical, contextual and grammatical information coordinates from the earliest possible stage. Research by Sedivy et al 1999 and especially Grodner & Sedivy (in press) on reference resolution indicates that pragmatic constraints that are similar to the ones we discuss in this paper facilitate disambiguation from the earliest stages of processing. On this view, one would have to explain why specifically SIs (as opposed to the phenomena under investigation in the studies above) are time consuming and not as local as expected. Further research could shed more light on this issue.

Acknowledgments

References

Altmann, G. & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *38*, 419-439.

Bezuidenhout, A. & Cutting, J.C. (2002). Literal meaning, minimal propositions, and pragmatic processing. *Journal of Pragmatics*, *34*, 433-456.

Bezuidenhout, A. & Morris, R. (2004). Implicature, Relevance and Default Inferences. In Noveck, I. & Sperber, D. (Eds.), *Experimental Pragmatics*. Basingstoke: Palgrave Macmillan

Bott, L. & Noveck, I .A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*, 437-457

Breheny, R. (forthcoming). Some Scalar Implications Really Aren't Quantity Implicatures - But Some Are. *Proceedings of Sinn & Bedeutung 9*, University of Nijmegen.

Breheny, R., Katsos, N. & Williams, J. in press) Are Generalised Scalar Implicatures Generated by Default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition*

Carston, R. (1998). Informativeness, relevance and scalar implicature. In R. Carston & S. Uchida (Eds.), *Relevance Theory: Applications and Implications*. Amsterdam: John Benjamins.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena and the syntax/pragmatic interface. In A. Beletti (Ed.), *Structures and Beyond*. Oxford University Press.

Ferreira, F. & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348-358.

Geurts, B. (1998). Scalars. In P. Ludewig and B. Geurts (Eds.), *Lexikalische Semantik*. Tubingen: Gunter Narr.

Grodner, D. & Sedivy, J. (in press) The effect of speaker-specific information on pragmatic inferences. In N. Pearlmutter & E. Gibson (Eds.), *The Processing and Acquisition of Reference.* MIT Press: Cambridge, MA

Huang, Y.T. & J. Snedeker (2005) What exactly do numerals mean? Paper presented at *Experimental Pragmatics*. Cambridge, UK.

Levinson, S. (2000). *Presumptive Meanings*. Cambridge, MA.: MIT Press.

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

Papafragou, A. & Musolino J. (2003). Scalar implicatures: experiments at the semantics/pragmatics interface. *Cognition*, *86*, 253-282.

Recanati, F. (2003). *Literal Meaning*. Cambridge: Cambridge University Press

Sedivy, J., Tanenhaus, M., Chambers, C., & Carlson, G. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109-147.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Spivey-Knowlton, M. & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*, 227-267.

Storto, G. & Tanenhaus, M. K. (2005) Are scalar implicatures computed online? *Proceedings of SuB 9*: 431-445.