

Pragmatics: From Theory to Experiment and Back Again

Napoleon Katsos* and Chris Cummins

Research Centre for English and Applied Linguistics, University of Cambridge

Abstract

This article reviews some cases in which the collaboration of theoretical pragmatics and psychologists of language has been most fruitful for all parties. Linguists have benefited from experimental data confirming the psychological validity of their observations and providing theory-critical evidence in cases beyond the reach of reflective intuition, whereas psychologists have benefited from having a wealth of linguistic phenomena to study as well as multiple theories furnished by semantics and pragmatics. Focusing on the much-discussed question of whether scalar implicature is a default inference, we aim to showcase the benefits of the experimental approach to pragmatics.

1. Introduction

Theoretically motivated empirical investigations of pragmatic and semantic aspects of meaning are not novel. However, in recent years, interest in experimental pragmatics has intensified, leading to a flourishing literature in both linguistic and psychological journals, comprising overview and position articles (Noveck and Reboul 2008; Sedivy 2007; i.a.), a series of meetings and conferences since 1998 and a dedicated website (<http://www.experimentalpragmatics.org>; Clark and Noveck 2005), and a European Science Foundation network grant (Euro-Xprag, see <http://www.euro-xprag.org>; Noveck et al. 2009), among others.

This article does not offer a historical review of previous and current investigations, although such a review might provide a useful sidelight on the evolution of psychological and linguistic approaches to the study of meaning (see Sperber and Noveck 2004). Rather, we present cases where interdisciplinary investigations between linguistics and psychology have proved mutually beneficial, whilst also highlighting problems to be overcome in this approach.

The discussion of several topics could serve this aim. Studies on reference, speech act, metaphor, irony, and so-called figurative language comprehension have produced a wealth of psycholinguistic findings and informed models of human cognition. Many of these investigations were directly motivated by, and have fed back into, the linguistic theory of the phenomena.

For instance, studies by Clark and colleagues on reference (see Clark and Bangerter 2004 for a review) indicate that the coordination condition proposed by Grice (1989) and Lewis (1969) is insufficient, and that reference assignment requires more collaboration between speaker and hearer. This work has significance not only in a linguistic context but also to the psychology of language processing and Theory of Mind, the ability to take into account interlocutors' epistemic state (see Hanna et al. 2003; Barr and Keysar 2007; i.a.). On another front, research on figurative language by Gibbs, Glucksberg and colleagues, among others, has produced evidence against the pervasive

notion in –psycholinguistics that pragmatics is secondary to syntax and semantics (see chapters 2 and 3 in Noveck and Sperber 2004 for an overview). Specifically, studies in metaphor comprehension have shown that apparently pragmatic processes of the type suggested by Grice (1989) and Searle (1979) are not costly, nor effortful, but exhibit the automaticity customarily associated with modular processes (see also Giora 2003).

However, in this article we shall focus on *scalar implicature*, a topic that has generated much of the most recent literature. This phenomenon is particularly amenable to empirical investigation for several reasons. First, there is widespread agreement between theoretical accounts of scalar implicature about the interpretation of scalar terms (at least in the most straightforward cases). The predictions that could differentiate between competing accounts are often too fine-grained to be tested with the traditional tools of the theoretical linguist, reflective introspection and intuition. Empirical investigation is therefore a more suitable testing procedure, as argued by Chierchia (2004), Levinson (2000) and Sperber and Wilson (1986/1995), among others. Furthermore, the competing theoretical accounts of implicature have achieved a considerable level of conceptual clarity and precision, making it possible to generate clear predictions about processing and acquisition from the linguistic claims about the nature and availability of scalar implicature. Finally, the proliferation of research on scalar implicatures reflects the importance of this topic to determining the nature of the semantics–pragmatics interface, as well as the key questions in psycholinguistics. Issues such as the locus of implicature, the role of context and the pragmatic intrusion on truth conditions have repercussions for the organisation of the whole semantic and pragmatic system, whereas issues such as the time course of integrating contextual information and its interaction with grammatically encoded meaning are broadly relevant to psycholinguistics.

This article focuses in particular on the issue of whether scalar implicatures are default inferences. Despite this comparatively narrow remit, we hope to draw some wider conclusions about how linguistic claims can be associated with experimental findings, and which methodologies can be used to address such claims. However, we first turn our attention to the linguistic phenomena.

2. *Scalar Implicature and the Generalised-Particularised Distinction*

Consider the following conversation:

- (1) The Bride: Have you seen Tommy?
 Bill: Big guy in the tux?
 The Bride: Yes.
 Bill: Then I saw him. I like his hair.
 The Bride: You promised you'd be nice!

(Kill Bill, Volume 2; script by Quentin Tarantino;
 Uma Thurman as the Bride; David Carradine as Bill)

Let us further assume that the following inferences can be drawn from Bill's last utterance:¹

- (2) a. Bill doesn't love, adore or worship Tommy's hair.
 b. Bill doesn't like anything else about Tommy besides his hair.
 c. Bill doesn't approve of Tommy as a groom for the Bride.

According to Grice (1989), (2a), (2b) and (2c) are cases of conversational implicature. These pragmatic inferences arise because of certain conventions of rational behaviour: for example, because the interlocutors assume one another to be cooperative and to obey certain rules regulating the quantity and relevance of information that ought to be exchanged.

In all versions of current pragmatic theory, the derivation of (2c) would be (more or less) straightforwardly accounted for by some version of a maxim of relation or relevance. However, the precise derivation of (2a) and (2b) has been the subject of much controversy. In the early neo-Gricean tradition, and particularly by Horn (1972, 1984) i.a.), it was noted that these inferences rely upon the existence of a scale of informativeness. For (2a), the scale of informativeness consists of verbs of liking, <like, love, adore, worship ...>. Terms of the right of this scale are informationally stronger than terms on the left, as the former entail the latter, but not vice versa. For (2b), the situation is more complex. An entailment scale could consist of things that one could like about a prospective groom: <{hair}, {personality}, ..., {hair, personality}, {hair, personality, ...}, ...>. Alternatively, the scale for (2b) could involve a person's characteristics, ranked in terms of noteworthiness, for example, <hair, ..., physique, ..., personality, ...>. In this case there is no entailment relation between the terms of the scale, but we can see how the terms could be ranked according to this (subjective) criterion.

However the scale is formed, the reasoning that derives (2a) and (2b) goes approximately as follows. The interlocutors are interested in establishing what Bill thinks of Tommy. Hence, any information on that point is relevant to the conversation. Now, the speaker said 'I like [Tommy's] hair'. If the speaker wanted to say something more informative, he would have done so. Assuming that he is cooperative and knows what he is talking about (valid as he is expressing his own views), his failure to say anything more informative implies that he did not want to. That is, the fact that Bill only said he *liked* Tommy's hair implies that he did not feel any more strongly about it, and the fact that Bill only said he liked Tommy's *hair* implies that he did not like anything else about Tommy, which arguably justifies the Bride's reproach.

As these inferences are conversational in nature (rather than grammatically encoded), and because the derivation relies upon the existence of informativeness scales, they are traditionally known as *scalar implicatures*. However, within this category, there may be important differences, for instance between (2a) and (2b). Consider the utterance 'I like Tommy's hair' occurring out of the blue, without any conversational context. Arguably, inference (2b) never occurs in this case, whereas (2a) still seems to be available. This is the difference that Grice (1989:37) himself had in mind when he introduced the distinction between Particularised and Generalised Conversational Implicatures (henceforth *PCIs* and *GCI*s):

[PCIs] are cases in which an implicature is carried by saying that *p* on a particular occasion in virtue of special features of the context, cases in which there is no room for the idea that an implicature is normally carried by saying that *p*. But there are cases of generalized conversational implicature. Sometimes one can say that the use of a certain form of words in an utterance would normally (in the absence of special circumstances) carry such-and-such an implicature or type of implicature.

Although the PCI–GCI distinction was explicitly introduced by Grice himself, it should be emphasised that, for Grice (1989:39–40), calculating an implicature of either type 'is to calculate what has to be supposed in order to preserve the supposition that the Cooperative Principle is being observed'. Thus, both PCIs and GCIs satisfy the criteria for conversational implicature status; for example, they are defeasible, calculable, indeterminate and non-detachable (see Horn 1984; Levinson 1983; i.a.).

It has generally been agreed that conversational implicatures ought to have at least some of these properties (Sadock 1978), and there is by now ample evidence that scalar implicatures (henceforth SIs) do. For example, with regard to defeasibility, it is expected that SIs can be cancelled without giving rise to strong contradiction. Katsos (2007; ch. 3) reports several studies where participants are presented with short discourses with an utterance containing a scalar expression followed by an utterance that contradicts either the SI or a semantic entailment of that expression. It was consistently found that the continuation contradicting the SI was rated as more coherent than that which contradicted the entailment (although not perfectly coherent).

Moreover, there is evidence that SIs are indeed context-dependent, and are not generated in conversational contexts where they do not contribute to the participants achieving the discourse goal. As Horn (1984:13) notes, '[A scalar] ... implicatum ... may be ... cancelled implicitly (by establishing the appropriate contexts, in which all that is relevant, or can be known, is the lower-bound)' (see also Sperber and Wilson 1986/1995:276ff and Levinson 2000:51ff for similar comments). Let us call contexts in which 'all that is relevant ... is the lower bound' *lower-bound* (LB) contexts, and contexts in which pragmatic inference is required to achieve the discourse goal *upper-bound* (UB) contexts. Then, in the following example (adapted from Levinson 2000), (3a) is an UB context and (3b) an LB context.

- (3) a. Were all their identity documents forgeries?
- b. Is there any evidence against them?
- c. Some of their identity documents are forgeries.

In answer to (3a), the speaker of (3c) implicates that not all of their identity documents were forgeries. This is a classical scalar implicature: 'all' is more informative than 'some' in this proposition, and if it were true to say 'all', the speaker would have said so, because it would have been relevant [given (3a)]. Hence, assuming that the speaker has the knowledge to make the stronger statement, it follows that (3c) gives rise to the aforementioned SI. However, the same utterance (3c), when used as a response to (3b), could receive a different interpretation. The discourse goal set by (3b) – to establish whether there is any incriminating evidence – is satisfactorily addressed by the semantic meaning of (3c), which answers (3b) in the affirmative. In such a context, it is irrelevant whether the strengthened alternative proposition containing 'all' is true or not, and hence the SI need not be derived. Katsos et al. (2005) report studies where participants who are asked to interpret scalar expressions (in pen-and-paper tasks) are indeed sensitive to the difference between implicature-generating UB contexts and implicature-free LB contexts.

There is also evidence that, in accordance with the Gricean account, interlocutors modify their expectations of informativeness according to each other's epistemic state. Chierchia et al. (1998), Musolino and Lidz (2006) and Katsos and Bishop (2008) have demonstrated that comprehenders only generate SIs in conditions where their speaker is known to have all the relevant information. For instance, given a scenario in which 'all' is applicable, a speaker's description of that scenario using the quantifier 'some' is consistently rejected on the grounds of underinformativeness. However, if the speaker is making a guess about an action that is yet to take place, or a bet about an action that has happened unseen, then participants are significantly more willing to accept 'some' in the scenarios where 'all' applies, demonstrating sensitivity to whether the speaker is in a position to make a more informative statement or not (Katsos and Bishop 2008).

From this brief review, we can see that the predictions common to all accounts of SI – that such inferences are defeasible, persist only when relevant to the discourse goal,

and take into account the epistemic state – are empirically supported. This allows us to proceed to the central issue of this article, namely the default or contextual nature of scalar implicature.

3. Positions on Defaultness

In the derivation of generalised implicatures, the concept of a lexical scale of informativeness is crucial. Strong constraints have been proposed on what constitutes a lexical scale: terms must be equally lexicalised, of the same form class, of a similar register and about the same semantic field (Horn 1984, 2004; see also Atlas and Levinson 1981; Levinson 1983, ch. 3, 2000:79ff). Lexical scales meeting these criteria are formed by quantifiers, connectives and modals, as in (4a) to (4c), as well as gradable adjectives (4d/e), articles (4f; see Hawkins 1991) and numerals (4g), among others.

- (4) a. <some, many, most, all>, <sometimes, always>, <something, everything>
- b. <or, and>
- c. <possible, certain>, <may, must>
- d. <warm, hot>, <cool, cold>
- e. <like, love, adore>, <dislike, hate>
- f. <a, the>
- g. <one, two, ...>

Alongside this constrained set of lexical scales, there is a practically unlimited range of scales that can be formed in an *ad hoc* fashion for specific conversational settings. This contrast is exhibited by (2a) and (2b) in Section 2: the former is context-independent because it relies on a lexical scale, whereas the latter is context-dependent, because it relies on an *ad hoc* scale that has only been introduced in a specific context. Removing or changing the context stops the generation of the implicature. This is captured by Horn (2004:3, emphasis in the original), who observes that ‘it is apparent that some implicatures are induced **only** in a special context [=PCIs] ... while others go through **unless** a special context is present [=GCIs]’.

Emphasising this distinction, Levinson (2000) and Chierchia (2004) explicitly argue that context is not at all necessary for deriving GCIs. They argue that GCIs are derived by the application of default principles that generate preferred, albeit potentially defeasible, inferences. Contextual factors may then cause these to be cancelled (e.g. if the speaker is identified as not fully cooperative or not in a position to make a stronger statement). In Levinson’s (2000) account, GCIs are derived by specialised pragmatic rules that have evolved to maximise the speed of communication. Chierchia (2004:57) bases his radical proposal ‘on the idea of adding [SIs] in as soon as possible, as part of the computational system of grammar, rather than as part of the interaction of such a system with other extra-grammatical modules’. In either account, PCIs like (2b) and (2c) would be derived by a distinct context-sensitive Gricean pragmatic system.

In contrast to these *default*² approaches, other accounts collapse the PCI–GCI distinction and maintain that (i) the same pragmatic principles are involved in generating PCIs and GCIs, and (ii) certain contextual conditions are required for the derivation of both kinds of implicatures. For example, Hirschberg (1991) provides an elaborate survey of lexical scales and ‘posets’, partially ordered sets of terms including a wide range of relations (whole/part, type/subtype, etc.) and scales based on contextually salient contrastive relations in a non-entailment ordering such as the one in (2b). In this account, lexical and logical scales are perhaps the most salient kind of scale, but nevertheless unexceptional.

Although radically different in the specifics of how implicatures are derived, relevance theory (Sperber and Wilson 1986/1995; see also Carston 1998, 2002; i.a.) also undermines the GCI–PCI distinction. According to this theory, all implicatures are derived by the same strategy, and contextual assumptions are assumed to be equally integral to deriving any type of pragmatic inference. Thus, if an SI is not contextually relevant, it will not be generated. As the relevance theory accounts, like Hirschberg’s account and other accounts such as Geurts (1998), assume a single context-dependent pragmatic system to be responsible for generating all SIs, we will refer to these as *context-driven* accounts.

4. Theory-Critical Predictions

Let us outline the issues where default and context-driven accounts make contrasting predictions. First, the accounts envisage different roles for discourse context. Context-driven accounts predict that all implicatures are generated only when relevant to the discourse goal, whereas default accounts predict that this only holds for PCIs, with GCIs also being generated in ‘neutral’ contexts where it cannot be established whether the implicature is relevant or not. In contexts where the implicature is discourse-irrelevant, default accounts – unlike context-driven accounts – further predict that the implicature is first generated and then cancelled.

Second, default accounts predict that GCIs are more easily generated than PCIs, for reasons of processing economy. According to Levinson’s account, default heuristics have developed to maximise the speed of communication, hence generating a GCI should be rapid and extra effort could be required when cancelling them. In contrast, relevance theory-inspired accounts tend to predict that generating any pragmatic inference is costly in terms of processing resources (but they do not normally need to be cancelled, hence not giving rise to additional costs – see Bott and Noveck 2004). However, these predictions are not necessarily shared by all accounts in their respective categories. For example, there is no stipulation in Chierchia’s default account that implicature generation or cancellation should be easy or slow, respectively, merely that SI generation is context-independent and therefore occasionally leads to cancellation and backtracking.

Theories of SI also differ in whether they assume local or global generation of implicatures, and whether these implicatures intrude upon truth conditions. For now, we leave aside these dimensions of variation and consider the experimental evidence available to address the question of default versus context-driven generation of SIs in terms of the role of context and the time course of generating SIs.

5. Experimental Investigations

Carston, Levinson, Chierchia and Sperber and Wilson are among those who consider that their accounts should enjoy psycholinguistic as well as linguistic validity, and that their predictions about the availability and generation of SIs should be empirically (dis)confirmable (cf. Levinson 2000:5, 81, 162ff, 370; Chierchia 2004:51, 68, 93; Carston and Powell 2006 for the relevance theory perspective). Relevance theory and Levinson’s account are actually motivated by cognitive considerations (principles arising from cognitive evolution and the psychology of sentence processing, respectively).

Furthermore, as we shall see in the following section, experimental investigation is necessary because several theory-critical predictions concerning SIs lie outside the reach of the traditional tools of the linguist – conceptual argumentation, introspection and intuition. At a time when it is widely acknowledged that syntax has run out of ‘cheap data’

and that a rapprochement of cognitive psychology with theoretical linguistics is highly desirable for both disciplines (see Ferreira 2005; Marantz 2005; i.a.), Chierchia (2004:46) makes precisely this point for SIs.

In the following subsections, we will review theoretically motivated empirical investigations of SIs, paying particular attention to their contribution to the theoretical debate.

5.1. THE ROLE OF CONTEXT

Context-driven and default accounts agree that SIs are ultimately available when relevant to the discourse goal and not available otherwise; however, they differ as to how this situation arises. Default accounts predict that the SI is first generated, regardless of context, then cancelled if it is contextually irrelevant, whereas context-driven accounts predict that the SI is not generated at all if it would not be relevant. Clearly intuition alone cannot settle this question, as we are concerned not with the ultimate interpretation but with the process through which it is derived. However, psycholinguistic investigations of the time course of scalar implicature can shed light on this.

More generally, a central debate in the area of sentence processing concerns whether there is an initial, encapsulated stage of structure assignment, or whether different categories of information (syntax, semantics, context, co-occurrence frequencies, etc.) interact and coordinate from the earliest possible stage (Spivey-Knowlton and Sedivy 1995; Tanenhaus et al. 1995; i.a.). The question of how SIs are generated could be seen as part of this debate.

According to the default approach, more processing time should be required when participants are processing scalar terms in LB contexts (where the SI is irrelevant) than in UB contexts (where it is relevant), because in the latter the inference must be generated and then cancelled. This additional processing should manifest itself as a delay in reading time, in self-paced reading studies. Context-driven accounts do not share this prediction, as no cancellation is required; in fact, accounts based on relevance theory (see Bott and Noveck 2004; i.a.) predict that reading time will be slower in UB contexts, because an additional inference is generated in these contexts. Thus, by investigating the reading times of scalar expressions in UB and LB contexts, we can gather evidence about the role of context as well as the automaticity of the inference.

In a series of self-paced reading experiments, Breheny et al. (2006; see also Katsos et al. 2005; Katsos 2008) investigated the reading time of scalar expressions in such contexts. For example, for the existential quantifier ‘some of the Fs’, Breheny et al. (2006) constructed UB and LB conditions, by manipulating whether ‘all of the Fs’ was relevant to the discourse (5a/c). They included a control condition with ‘only’ (5b) that explicitly encoded the SI.³

- (5) a. (UB) Mary asked John whether he intended to host all his relatives in his tiny apartment. John replied that he intended to host some of his relatives.
The rest would stay in a nearby hotel.
- b. (Control) Mary asked John whether he intended to host all his relatives in his tiny apartment. John replied that he intended to host only some of his relatives. The rest would stay in a nearby hotel.
- c. (LB) Mary was surprised to see John cleaning his apartment and asked the reason why. John told her that he intended to host some of his relatives.
The rest would stay in a nearby hotel.

The reading time was measured on the scalar term itself (some of his relatives) and on a dependent phrase (the rest) whose interpretation is facilitated by the SI. If the participants

interpret 'some' as 'some but not all', then the reference set for 'the rest' is already available when they read the dependent phrase; if they do not, then extra processing time is required to interpret who 'the rest' refers to.

As predicted by all accounts, the dependent phrase was read faster in the UB and control conditions than in the LB. This indicates that the SI was ultimately generated in UB and not in LB, indicating that the conditions were appropriately constructed. To answer the theory-critical question – whether the SI is generated and then cancelled in the LB condition – we have to look to the scalar expression itself. In this study (and in Katsos et al. 2005 for English), the trigger phrase 'some of the Fs' was read more slowly in UB than in LB contexts. This is hard to reconcile with the default account, as no reading time delay was evident in the LB condition. Thus, these findings show that contextual factors have a primary role in the process that generates SIs, not merely in cancelling them at a later stage. (For experimental studies that reach different conclusions, see Bezuidenhout and Cutting 2002; but see also the discussion of that paper in Breheny et al. 2006).

The second prediction on the role of context concerns 'neutral' situations in which the discourse goal cannot be clearly established. In these cases, default accounts predict generation of the SI, whereas context-driven accounts do not. Note that single sentence utterances are not necessarily neutral, as they can create their own context through a variety of presupposition triggers and information-structure triggers. For example, in (6), the NP 'Mary' is linguistically marked as focus, that is, the answer to an implied contextual question.

(6) It was Mary who ate shark-soup.

If the question 'who ate shark-soup?' is not already salient for the hearer, an utterance of (6) would trigger its accommodation into the context. Similarly, (7a) is more likely to relate to an issue like 'what happened to the consultants?' or 'who did the consultants meet with?' than is (7b).

- (7) a. Some of the consultants had a meeting with the director.
- b. The director had a meeting with some of the consultants.

This was controlled for in a further experiment by Breheny et al. (2006), conducted in Greek. Greek has a strong tendency for grammatical subjects in sentence-initial position to be construed as old information and post-verbal objects to be construed as unrelated to a contextual issue (see Philippaki-Warbuton 1987; i.a.) Breheny et al. created two-sentence discourses in Greek in which 'some of the Fs' was either a pre-verbal subject (8a) or post-verbal object (8b).

- (8) a. Some of the consultants had a meeting with the director.
 The rest did not manage to attend.
- b. The director had a meeting with some of the consultants.
 The rest did not manage to attend.

The target phrase ('the rest' in the foregoing examples) can be interpreted immediately but requires making a bridging inference that the collection referred to by 'the rest' is the rest of the consultants. If the SI 'not all the consultants' has been generated online in the first sentence, this inference will be facilitated. Given the assumptions about the effect of the word order on accommodating contextual issues, context-driven accounts predict that this will affect the reading time for 'the rest'. If there is no accommodation of an appropriate contextual issue, such as 'what happened with the consultants?', there is no reason for the SI to become available in the neutral context. Thus, reading time should only be

facilitated by the trigger being in sentence-initial position. In contrast, default accounts predict that the implicature should be generated online regardless of where the implicature trigger is, and hence that sentence position should not affect reading time.

Additionally, control conditions were created in which the inference was explicitly marked with 'only', as in (8c) and (8d), which was expected to result in maximally short reading times.

- (8) c. Only some of the consultants had a meeting with the director.
The rest did not manage to attend.
- d. The director had a meeting with only some of the consultants.
The rest did not manage to attend.

The results showed that 'the rest' was read equally quickly in the two cases where 'some' was sentence-initial (8a/c): this suggests that the SI was generated in (8a) and the referent set was as readily available as when introduced by 'only'. However, the reading time for 'the rest' was slower in (8b) than in (8d), suggesting that when 'some of the consultants' was sentence-final, the contrast set was not as readily available as a referent for 'the rest'. Since, as we have argued, (8b) is a neutral context for a scalar expression, these findings suggest that the predictions of default accounts are not upheld.

5.2. THE TIME COURSE OF SCALAR IMPLICATURE

In a series of studies in French, Noveck and colleagues (Noveck 2001; Noveck and Posada 2003; Bott and Noveck 2004) investigated the time course of SI generation. They presented participants with utterances containing scalar terms which are logically true but underinformative given readily available world knowledge, such as in (9).

- (9) Some elephants have trunks.

Participants were asked to give a truth-value judgement on such sentences. These would be 'false' if, as Noveck argues, they generated the SI (some but not all), or 'true' if they accessed the plain meaning of the quantifier 'some'. In experiment 3 of Noveck (2001), it was found that about 60% of such utterances are rejected. Noveck and Posada (2003) measured response times to utterances containing a scalar term after the whole utterance had been processed, and found that the time taken by the participants who were instructed to answer 'false' was significantly longer than those who were instructed to answer 'true'. Bott and Noveck (2004) replicated these findings both with respect to the percentage of negative responses and the time course of the two responses, even when substituting the response 'true'/'false' (which might bias for logical interpretations) for 'I agree'/'I disagree' (which might be more relevant to felicity than to logical truth-value judgements).

Bott and Noveck (2004) also excluded the possibility that the time difference between the two responses was because of the difference in response type (positive/negative). They showed that no time difference was evident in the corresponding control conditions (accepting or rejecting utterances that are patently true and informative, or patently false). They also showed that the same pattern of results obtained in a more complex experiment in which the utterances were narrated by a character (thus allowing answers of both types to have the same response type). This pattern persisted when all instructions were removed. In a fourth experiment, they found that the number of responses based on the SI increased as the participants were permitted more time to respond.

The studies by Noveck and colleagues yielded evidence that is relevant to two predictions. First, they showed that adult native speakers do not always generate implicatures in

simple declarative sentences, as predicted by default accounts. Second, they show that generating an SI is not an automatic process, but rather an effortful one. If, as according to default accounts, generating an SI were automatic and cancelling it required extra resources, then the acceptance of underinformative utterances ought to take longer than the responses where they reject them; moreover, when participants are given less time to respond, then the response pattern that should increase is the acceptance, rather than the rejection, of underinformative utterances. The findings of Noveck and colleagues have been replicated in English using the same methodology (De Neys and Schaeken 2007; but see also Feeney et al. 2004), and there is also evidence from studies using the visual world eye-tracking paradigm that corroborates the claim that generating SIs is indeed a time-consuming process (Huang and Snedeker 2009).

6. Overview and Outlook

In the previous section, we reviewed empirical investigations that have been motivated by the debate between default and context-driven accounts of SI. We saw that there is mounting evidence that SIs are only generated in contexts where the inference is relevant to the discourse goal; in other contexts, it is not the case that they have first been generated and then cancelled. These findings are predicted by context-driven accounts and cannot be accommodated by default accounts. Moreover, generating SIs was found to be a time-consuming rather than a fast process, which also tends to cohere with context-driven rather than default proposals, although claims about the time course of the inference need not be an integral part of default, or context-driven, accounts. The mounting evidence in favour of context-driven accounts that consider pragmatic inferences as extra-grammatical and effortful lends support to accounts similar to the original Gricean proposal, but may also accept other interpretations in future work.

With regard to methodology, we have discussed several paradigms for investigating different aspects of the SI generation process. Pen-and-paper studies are perfectly adequate when we are interested in the eventual interpretation of the expression being studied (e.g. defeasibility of SIs, relevance of epistemic state). However, self-paced reading, eye-tracking and reaction time studies are called for when we are investigating the process through which the eventual interpretation is generated.

In all these cases, the relation between conceptual clarity and successful experimentation can be seen as a two-way street; clear predictions at the theoretical level render a topic straightforwardly amenable to empirical study, whereas the drive for empirical investigations can underscore the need for clarity at the theoretical level.

We conclude this article by highlighting further outstanding questions in the debate on SI generation, and discuss how the collaboration of linguists and psychologists could be fruitful in resolving these.

6.1. WHY SHOULD SCALAR IMPLICATURE BE EFFORTFUL?

In Section 5.2, we saw that generating an SI is an effortful, costly process. But why should SIs be costly? There is by now ample evidence that pragmatic inferences *per se* are not necessarily costly in processing terms. For example, work on metaphor comprehension and indirect speech acts (Gibbs 1986; McElree and Nordlie 1999; Glucksberg 2003; Gibbs and Tendahl 2006; i.a.) has shown that drawing the pragmatically appropriate interpretation can be automatic and not involve extra cost when the appropriate interpretation is clear-cut.

So why should SIs *specifically* be costly? Multiple suggestions are available: SIs are meta-linguistic inferences requiring the addressee to consider the possible alternative utterances available to the speaker; the negative component of the SI ('p or q but *not* both' for disjunction, 'some but *not* all' for the existential quantifier) adds to processing cost; or, as proposed by Noveck and colleagues, inferences that contribute to making an utterance more informative are generally costly, a cost which is balanced by the gain in informativeness.

Still another proposal is that the SI is not intrinsically costly, but gives rise to costly operations on the mental discourse model. When participants interpret 'Some of the Fs G-ed' without an SI, they introduce the set of Fs who G-ed into their discourse model. When they interpret it with an SI, they have to introduce the set of Fs and also partition it into two non-empty subsets (those who G-ed and those who did not G). The suggestion that the introduction of contrast sets, rather than the SI, induces a processing cost is compatible with two pieces of evidence. Grodner et al. (2005) showed that introducing contrast sets through restrictive relative clauses gives rise to processing costs. It has also been shown that, once a contrast set is established, inferential processes dependent on the second maxim of quantity (requesting speakers not to be over-informative) are initiated at the earliest possible stage, without manifesting any processing cost (Sedivy et al. 1999; Sedivy 2001; i.a.). Further research along these lines could shed light on the time course of SI processing, and whether this is something inherent to scalar reasoning.

6.2. ARE THERE NO DEFAULTS?

The online studies reported in Section 5.2 concerns the role of relevance of the SI to the discourse context. A distinct question concerns whether global conversational assumptions about cooperativity, epistemic state and so on are constructed for each communicative exchange, or whether they are introduced across the board by general principles of communicative behaviour. In the experiments where participants generated SIs, they must have assumed that the speaker was cooperative and had adequate knowledge of the topic under discussion, despite there being no explicit information to that effect.

One possibility is that people, at least in certain communicative settings with fixed social roles, use stereotypical assumptions on whether their interlocutors are cooperative and well-informed on the issues they discuss. This could indicate that there are 'defaults' at the social level or merely indicate that the assumptions are context-dependent and constructed from scratch for each communicative exchange.

Testing these assumptions would be another straightforward extension of the research presented in this overview, and would further broaden the scope of the debate by introducing wider sociological considerations on human interaction.

Short Biographies

Napoleon Katsos is interested in semantics and pragmatics, and particularly in quantification, conversational implicature and informativeness. He takes an interdisciplinary perspective and uses techniques and models developed in experimental psychology to address topics in theoretical linguistics; and vice versa, he aims to draw upon theoretical linguistics to provide fine-grained hypotheses and explanations that are relevant to psychologists of language. He obtained his PhD from the Research Centre for English and Applied Linguistics, University of Cambridge, on implicature and sentence processing. During his post-doctoral research at the Oxford Study for Children's Communication Impairments

he investigated the acquisition of quantification by young children. He has since returned to the Research Centre for English and Applied Linguistics working on these topics with a particular emphasis on what atypically developing children and cross-linguistic comparisons can reveal about the architecture of the linguistic system.

Chris Cummins is a graduate student in Experimental Pragmatics at the Research Centre for English and Applied Linguistics, University of Cambridge. His principal research interests concern the interpretation and use of numerically quantified expressions, and the relation of this to underlying cognitive considerations. He also conducts experimental work on scalar implicature. Other interests include: connectionist modelling of semantic acquisition and skilled reading, statistical learning and techniques for linguistic data analysis. He is supported by a University of Cambridge Domestic Research Studentship.

Notes

* Correspondence address: Napoleon Katsos, RCEAL, 9 West Road, CB3 9 DP, Cambridge, UK.
E-mail: nk248@cam.ac.uk

¹ These inferences are licensed on the grounds that Bill's last utterance is not ironic, that is, that he does mean he likes Tommy's hair. This is clear in the film, in that there is nothing wrong with Tommy's hair, whereas Bill has other reasons to disapprove of Tommy.

² It should be emphasised that by 'default' accounts we refer to approaches that propose defaultness at a *pragmatic* level, not on a conceptual level (e.g. Jaszczolt 2005).

³ The original experiment was conducted in Greek.

Works Cited

- Atlas, J., and S. Levinson. 1981. It-clefts, informativeness and logical form. *Radical pragmatics*, ed. by P. Cole, 1–61. New York: Academic Press.
- Barr, D. J., and B. Keysar. 2007. Perspective taking and the coordination of meaning in language use. *Handbook of psycholinguistics*, 2nd edn., ed. by M. J. Traxler and M. A. Gernsbacher, 901–38. New York: Academic Press.
- Bezuidenhout, A., and J. C. Cutting. 2002. Literal meaning, minimal propositions, and pragmatic processing. *Journal of Pragmatics* 34. 433–56.
- Bott, L., and I. A. Noveck. 2004. Some utterances are underinformative: the onset and time course of scalar inferences. *Journal of Memory and Language* 51. 437–57.
- Breheny, R., N. Katsos, and J. Williams. 2006. Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences *Cognition* 100(3). 434–63.
- Carston, R. 1998. Informativeness, relevance and scalar implicature. *Relevance theory: applications and implications*, ed. by R. Carston and S. Uchida, 179–236. Amsterdam: John Benjamins.
- . 2002. *Thoughts and utterances: the pragmatics of explicit communication*. Oxford: Blackwell.
- , and G. Powell. 2006. Relevance theory – new directions and developments. *Oxford handbook of philosophy of language*, ed. by E. Lepore and B. Smith, 341–60. Oxford: Oxford University Press.
- Chierchia, G. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. *Structures and beyond*, ed. by A. Belletti, 39–103. Oxford: Oxford University Press.
- , S. Crain, M. T. Guasti, and R. Thornton. 1998. "Some" and "or": a study on the emergence of logical form. *Proceedings from the 22nd Annual BUCLD*, ed. by Annabel Greenhill, Mary Hughes, Heather Littlefield, and Hugh Walsh, 97–108. Somerville, MA: Cascadia Press.
- Clark, H. H., and A. Bangerter. 2004. Changing conceptions of reference. *Experimental pragmatics*, ed. by I. Noveck and D. Sperber, 25–49. Basingstoke, England: Palgrave Macmillan.
- Clark, B., and I. A. Noveck. 2005. 11th Sep. 2009 <<http://www.experimentalpragmatics.org/>>.
- De Neys, W., and W. Schaeken. 2007. When people are more logical under cognitive load: dual task impact on scalar implicature. *Experimental Psychology* 54. 128–33.
- Feeney, A., S. Scafton, A. Duckworth, and S. J. Handley. 2004. The story of *some*: everyday pragmatic inference by children and adults. *Canadian Journal of Experimental Psychology* 58(2). 121–32.
- Ferreira, F. 2005. Psycholinguistics, formal grammars, and cognitive science. *Linguistic Review* 22. 365–80.

- Geurts, B. 1998. Scalars. *Lexikalische Semantik aus kognitiver Sicht*, ed. by P. Ludewig and B. Geurts, 95–118. Tübingen: Gunter Narr.
- Gibbs, R. 1986. *The poetics of mind*. Cambridge: Cambridge University Press.
- , and M. Tendahl. 2006. Cognitive effort and effects in metaphor comprehension: relevance theory and psycholinguistics. *Mind & Language* 21(3). 379–403.
- Giora, R. 2003. *On our mind: salience, context, and figurative language*. New York: Oxford University Press.
- Glucksberg, S. 2003. The psycholinguistics of metaphor. *Trends in Cognitive Science* 7. 92–6.
- Grice, H. P. 1989. *Studies in the way of words*. Cambridge, MA: Harvard University Press.
- Grodner, D., E. Gibson, and D. Watson. 2005. The influence of contextual contrast on syntactic processing: evidence for strong-interaction in sentence comprehension. *Cognition* 95. 275–96.
- Hanna, J. E., M. K. Tanenhaus, and J. C. Trueswell. 2003. The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language* 49. 43–61.
- Hawkins, J. 1991. On (in)definite articles: implicatures and (un)grammaticality prediction. *Journal of Linguistics* 27(2). 405–42.
- Hirschberg, J. 1991. *A theory of scalar implicature*. New York: Garland.
- Horn, L. 1972. On the semantic properties of logical operators in English. UCLA Dissertation, distributed by Indiana University Linguistics Club, 1976.
- , 1984. Toward a new taxonomy for pragmatic inference. *Meaning, form and use in context: linguistic applications*, Proceedings of GURT'84, ed. by D. Schiffrin, 11–42. Washington, DC: Georgetown University Press.
- , 2004. Implicature. *Handbook of pragmatics*, ed. by L. Horn and G. Ward. 3–28. Malden, MA: Blackwell Publishers.
- Huang, Y., and J. Snedeker. 2009. Online interpretation of scalar quantifiers: insight into the semantics–pragmatics interface. *Cognitive Psychology* 58(3). 376–415.
- Jaszczolt, K. 2005. *Default semantics: foundations of a compositional theory of acts of communication*. Oxford: Oxford University Press.
- Katsos, N. 2007. *Experimental investigations on the effects of structure and context on the generation of scalar implicatures*. PhD thesis, University of Cambridge.
- , 2008. The semantics/pragmatics interface from an experimental perspective: the case of scalar implicature. *Synthese* 165. 358–401.
- , and D. Bishop. 2008. The development of informativeness from a speaker's and comprehender's perspective. Paper presented in IASCL 2008, Edinburgh, UK.
- , R. Breheny, and J. Williams. 2005. The interaction of structural and contextual constraints during the on-line generation of scalar inferences. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, ed. by B. Bara, L. Barsalou and M. Bucciarelli, 1108–13. Mahwah, NJ: Erlbaum.
- Levinson, S. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- , 2000. *Presumptive meanings*. Cambridge, MA: MIT Press.
- Lewis, D. K. 1969. *Convention: a philosophical study*. Cambridge, MA: Harvard University Press.
- Marantz, A. 2005. Generative linguistics within the cognitive neuroscience of language. *Linguistic Review* 22. 429–45.
- McElree, B., and J. Nordlie. 1999. Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review* 6. 486–94.
- Musolino, J., and J. Lidz. 2006. Why children aren't universally successful with quantification. *Linguistics* 44(4). 817–52.
- Noveck, I. A. 2001. When children are more logical than adults. *Cognition* 86. 253–82.
- , B. Geurts, and U. Sauerland. 2009. 11th Sep. 2009 <<http://www.euro-xprag.org/>>.
- , and A. Posada. 2003. Characterizing the time course of an implicature: an evoked potentials study. *Brain and Language* 85. 203–10.
- , and A. Reboul. 2008. Experimental pragmatics: a Gricean turn in the study of language. *Trends in Cognitive Sciences* 12(11). 425–31.
- , and D. Sperber. 2004. *Experimental pragmatics*. Basingstoke: Palgrave Macmillan.
- Philippaki-Warbuton, I. 1987. The theory of empty categories and the pro-drop parameter in modern Greek. *Journal of Linguistics* 23. 289–318.
- Sadock, J. 1978. On testing for conversational implicature. *Syntax and semantics 9: pragmatics*, ed. by P. Cole, 281–97. New York: Academic Press.
- Searle, J. 1979. *Expression and meaning*. Cambridge: Cambridge University Press.
- Sedivy, J. 2001. Evidence for Gricean mechanisms in on-line language processing. Paper presented at the 14th Annual CUNY Conference on Human Sentence Processing, Philadelphia, PA.
- , 2007. Implicatures in real-time conversation: a view from language processing research. *Philosophy Compass* 2/3. 475–96.
- , M. Tanenhaus, C. Chambers, and G. Carlson. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71. 109–47.

- Sperber, D., and I. A. Noveck. 2004. Introduction to experimental pragmatics. *Experimental pragmatics*, ed. by I. A. Noveck and D. Sperber, 1–23. New York: Palgrave Macmillan.
- , and ———. 1986/1995. *Relevance: communication and cognition*. Oxford: Blackwell.
- Spivey-Knowlton, M., and J. Sedivy. 1995. Resolving attachment ambiguities with multiple constraints. *Cognition* 55. 227–67.
- Tanenhaus, M. K., M. J. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268. 1632–4.