On the suspension of implicatures

Anton Benz
Centre for General Linguistics, Berlin

Abstract

In this talk we apply the idea of preferential models from nonmonotonic semantics to game theoretic pragmatics in order to explain the mechanisms underlying the cancellation and suspension of implicatures. Preferential models represent assumptions about what is *normal.* We use them to represent assumptions about what are normal utterance situations. If the use of a form contradicts these assumptions about normality, then the addressee can conclude that the utterance situation is non-normal. This then can lead to the suspension of implicatures.

1 Introduction

The topic of this paper is an old one: that of suspension and cancellation of implicatures. The difference between these two forms of annulling an implicatures can be seen from the following examples:

(1) a. Cancellation: *Some, in fact all, of the boys came to the party.*
    b. Suspension: *Some, perhaps all, of the boys came to the party.*

In both cases, the scalar implicature from *some* to *some but not all* is cancelled. In case of cancellation proper, there is an explicit contradiction between the scalar implicature and the literal content of the utterance. In the case of suspension, the speaker only asserts the possibility of its negation (Horn 1989: pp. 234-235). The classical account is that of Gazdar (1979) who assumes that cancellation and suspension result from an incremental amplification of utterance interpretation. In this amplification, firstly all logical consequences are added to an utterance, secondly, if consistent, all clausal implicatures, thirdly all scalar implicatures, and finally all presuppositions. Hence, annulation of scalar implicatures is explained by a contradiction with the logical consequences, and suspension by a contradiction with the clausal implicatures of an utterance.

In this paper, we propose a game theoretic model which not only explains the cancellation and suspension of scalar implicatures but also of *relevance* implicatures. We discuss the following principled examples, which cannot be explained by a Gazdarian incremental account:

(2) a. A: *Does this job candidate speak Spanish?*
        i. *He speaks Portuguese.* +> He does not speak Spanish.
        ii. B: *I know he speaks Portuguese.* +> B does not know whether he speaks Spanish.
    b. A: *How did the students do in the exam?*
        i. B: *Some students passed.* +> Not many passed.
        ii. B: *I know that some students passed.* +> B does not know whether many passed.

The first example, (2ai), shows a relevance implicature. Hirschberg (1991) explained it as a generalised scalar implicature which results from a relevance scale. The suspensions in (2aii) and (2bii) cannot be explained by an incompatibility with a clausal implicature as *know* does not generate clausal implicatures (Gazdar 1979; Levinson 2000). Hence, an incremental account would predict that the (i) and (ii) versions lead to exactly the same implicatures.

From a logical point of view, implicatures are a kind of nonmonotonic inferences. A large class of nonmonotonic logics are concerned with inferences to what *normally* holds. For these logics, *normality* can be semantically defined by *preferential models* (Shoham 1987; Schlechta 2004). We borrow the idea of preferential models and apply it to a game theoretic model of implicatures as developed in Benz & van Rooij (2007).

Speaking very generally, a nonmonotonic inference $T \mid\!\sim R$ is valid in a nonmonotonic logic based on *preferential models* iff $M(T) \models R$ holds classically in a set $M(T)$ of preferred T models. $M(T)$ represents the assumptions about what normally holds if T is true. Let us again consider the standard example of a scalar implicature: *Some of the boys came to the party* (T) +> *Not all of the boys came to the party* (R). In the common explanation of this implicature, it must be assumed that the speaker knows how many of the boys came, i.e., $M(T)$ only contains possible worlds in which the speaker knows the actual world. If the expert condition is violated, then the implicature doesn't arise anymore, as can be seen from (1b). Accordingly, we will propose a game theoretic model in which a joint preference relation over a partition of all speaker types is defined such that the implicature of an utterance is calculated with respect to the minimal partition in which the utterance is licensed. The standard implicatures then follow from a preference for types with *expert* speakers. They are suspended if the utterance can only be *optimal* if the expert assumption is violated. We show that this model can explain the examples in (2).

2 The Standard Theory

In the standard theory (Gazdar 1979; Levinson 1983), suspension is explained by a conflict between scalar implicatures and *clausal* implicatures, which is resolved in favour of the latter. Clausal implicatures are a special case of quantity implicatures. Like scalar implicatures they result from a scale. They are not triggered by the occurrence of a weaker expression within a proposition, but by the use of a weaker operator which subordinates a proposition. The following table shows some examples:

a) stronger form	b) weaker form	c) implicature of weaker form
know A	believe A	$\Diamond A \wedge \Diamond\neg A$
necessarily A	possibly A	$\Diamond A \wedge \Diamond\neg A$
A and B	A or B	$\Diamond A \wedge \Diamond\neg A \wedge \Diamond B \wedge \Diamond\neg B$

From the assumption that the speaker follows the maxim of quantity, it follows that the use of the weaker form *believe* A implies that the speaker does not *know* A, hence, for all what the speaker knows, it is possible that A is false ($\Diamond\neg A$). Applied to our example (1b), it follows that *I believe that some of the boys came to the party* implicates that $\Diamond$ *all came* & $\Diamond$ *not all came.* And analogously it follows that *possibly all of the boys came* implicates that $\Diamond$ *all came* & $\Diamond$ *not all came.*

Scalar and clausal implicatures may get in conflict with each other. This conflict is resolved by the above mentioned incremental process which amplifies utterance meaning

(Gazdar 1979). The examples in (2) pose a principled problem for this account as *know,* being the strongest expression, does not generate clausal implicatures, and hence should not suspend any of the scalar implicatures.

Before turning to the critical examples, we first consider an explanation which derives scalar implicatures and their suspension directly from the assumption that the speaker follows the Gricean maxims. We write $A(X)$ for the sentence frames X *of the boys came to the party*' and X *students passed* with open quantifier position X; $A(\forall)$ and $A(\exists)$ means that *all* and *some* fill the open position, and $A(\exists \wedge \neg\forall)$ that *some but not all* occupies it; UA means that the speaker uttered A; and $\Box A$ means that the speaker is convinced that A. "*(Quality)*" says that the speaker can only utter what he believes to be true, and "*(Quantity)*" that he will always utter the strongest form he is convinced of:

Derivation of Implicature in Example 2b

1. $\Box A(\forall) \rightarrow \mathrm{U}A(\forall)$ *(Quantity)*
2. $\mathrm{U}A(\exists)$ (fact)
3. $\neg\Box A(\forall)$ (follows from lines 1 & 2)
4. $\Box A(\exists)$ (follows from 2 and *Quality*)
5. $\Box A(\neg\exists) \vee \Box A(\exists \wedge \neg\forall) \vee \Box A(\forall)$ *(Expert)*
6. $\Box A(\exists \wedge \neg\forall)$ (follows from lines 3., 4., and 5.)

Suspension of Implicature in Example (1b)

1. $\Box A(\exists) \wedge \Diamond A(\forall)$ (logical form of utterance and *Quality*)
2. $\Box A(\neg\exists) \vee \Box A(\exists \wedge \neg\forall) \vee \Box A(\forall)$ *(Expert)*
3. $\Box A(\forall)$ (follows from previous lines)
4. $\Box A(\forall) \rightarrow \mathrm{U}A(\forall)$ *(Quantity)*
5. *Contradiction* (because speaker did not utter $A(\forall)$)
6. $\neg$*(Expert)* $\equiv \Diamond\neg A(\neg\exists) \wedge \Diamond\neg A(\exists \wedge \neg\forall) \wedge \Diamond\neg A(\forall)$
7. $\Box A(\exists) \wedge \Diamond A(\neg\forall) \wedge \Diamond A(\forall)$ (from the first and the previous line)

These derivations are close to Grice' (1989) original account of implicatures. The *(Expert)* assumption says that the speaker knows whether A of *none, some but not all* or *all* is true. The suspension examples with *possibly* make it very clear that this assumption is necessary for deriving the standard scalar implicature. That it is often omitted indicates that it is an assumption that is tacitly made about speakers in *normal* situations. In the suspension example (1b), this assumption conflicts with the assumption that the speaker follows *(Quantity).* This conflict makes it necessary for the interpreter to decide which assumption to give up. With Grice, we assume that, in the face of a conflict, the interpreter holds on to the assumption that the speaker follows the maxims. We can see from the second derivation that giving up the *(Expert)* assumption not only suspends the scalar implicature $\Box A(\exists \wedge \neg\forall)$ but also explains the clausal implicature $\Diamond A(\neg\forall) \wedge \Diamond A(\forall)$. In the following section, we present a general framework that allows integrating this reasoning in a game theoretic model of implicatures.

3 The Game Theoretic Model

Before introducing a general model, we first provide an analysis of example (2a). To make the scenario more specific, we assume that the inquirer has to decide between two candidates a and b. Let $S(x)$ and $P(x)$ be the propositions 'x speaks Spanish' and 'x speaks Portuguese', and let $\Box X$ mean that the speaker is convinced that X. We represent the resulting 16 possible worlds on the left side of the table in figure (3). On the right side of the top row, 8 answers with low complexity are listed.

(3) *A model of example (2a)*

$S(a)$	$P(a)$	$S(b)$	$P(b)$	$S(a)$	$P(a)$	$\neg S(a)$	$\neg P(a)$	$\Box S(a)$	$\Box P(a)$	$\Box\neg S(a)$	$\Box\neg P(a)$
+	+	+	+	1	1	·	·	1	1	·	·
+	+	+	−	1	0	·	·	1	0	·	·
+	+	−	+	1	1	·	·	1	1	·	·
+	+	−	−	1	1	·	·	1	1	·	·
+	−	+	+	1	·	·	1	1	·	·	1
+	−	+	−	1	·	·	1	1	·	·	1
+	−	−	+	1	·	·	1	1	·	·	1
+	−	−	−	1	·	·	1	1	·	·	1
−	+	+	+	·	1	1	·	·	1	1	·
−	+	+	−	·	1	1	·	·	1	1	·
−	+	−	+	·	1	1	·	·	1	1	·
−	+	−	−	·	1	1	·	·	1	1	·
−	−	+	+	·	·	1	1	·	·	1	1
−	−	+	−	·	·	1	1	·	·	1	1
−	−	−	+	·	·	1	0	·	·	1	0
−	−	−	−	·	·	1	1	·	·	1	1

We assume that the inquirer knows the truth values of $S(b)$ and $P(b)$, and that the question in (2a) is about a. Furthermore, we assume that a job candidate with knowledge of Spanish is much preferred over a candidate without it. We interpret *much preferred* as meaning that if the inquirer has to choose between a candidate x of whom he knows that he speaks Spanish, and another candidate y of whom it is uncertain whether he speaks Spanish, he will decide for the candidate x. If the job candidates's knowledge of Spanish is equal, then the candidate who speaks Portuguese is preferred. The numbers 1 and 0 in the answer columns mean that the inquirer will (1) or will not (0) choose an optimal candidate given that he learns the answer and assumes all worlds to be equally probable. The dots mean that the respective answer is not admissible. For example in the second row, the inquirer knows that $S(b) \wedge \neg P(b)$. If he learns that $S(a)$, he will decide for candidate a as a may also know Portuguese with probability $\frac{1}{2}$. This is the correct choice (1). If he learns that $P(b)$, then he will decide for b as he knows that b speaks Spanish, and a Spanish speaking candidate is much preferred over a candidate who doesn't speak Spanish, which may be true of a with probability $\frac{1}{2}$. This is the wrong choice (0). In all rows we find that the boxed answers lead to the same results as the unboxed answers.

What are the optimal answers? Lets first assume that the speaker is an expert about a. He is not an expert about b, i.e., he knows the truth values of $S(a)$ and $P(a)$ but not of $S(b)$ and $P(b)$. The optimal answers in situations in which $S(a) \wedge P(a)$ holds are $S(a)$ and $\Box S(a)$. We write $\mathrm{Op}_{++} = \{S(a)\}$ for the set of un-boxed optimal answers. Analogously, we arrive at $\mathrm{Op}_{+-} = \{S(a), \neg P(a)\}$, $\mathrm{Op}_{-+} = \{P(a), \neg S(a)\}$, and $\mathrm{Op}_{--} = \{\neg S(a)\}$. Hence, from the fact that the speaker only produces optimal answers, the inquirer can infer that answer $P(a)$ implicates that $\neg S(a)$, and that answer $\neg P(a)$ implicates that $S(a)$. The optimal answers $S(a)$ and $\neg S(a)$ do not lead to additional implicatures.

Before considering the non-expert case, we present the game theoretic framework in which this reasoning can be carried out.

Our basic structures representing utterance situations are *interpreted support problems,* i.e., tuples $\sigma = \langle \Omega, P_S, P_H, \mathcal{F}, \mathcal{A}, u, c, [\![\,.\,]\!] \rangle$ consisting of (1) a finite set of possible worlds Ω, (2) probability distributions P_S for the speaker and P_H for the inquirer, or hearer, (3) a finite set of forms $\mathcal{F}$ from which the speaker chooses his answer, (4) a set $\mathcal{A}$ of hearer actions, (5) a joint payoff function u, (6) a function c measuring the complexity of forms, and (7) an interpretation function $[\![.]\!]$ which maps forms to subsets of

Ω. If, in situation v, the speaker chooses form F and the hearer action a, then the joint payoff is $u(v,a) - c(F)$. The costs $c(F)$ of forms are assumed to be nominal, i.e., positive but very small. We represent Grice' maxim of quality by the assumption that the speaker can only choose answers which he believes to be true, i.e., answers from the set $Adm_\sigma = \{F \in \mathcal{F} \mid P_S([\![F]\!]) = 1\}$. We further assume that the beliefs of the speaker and the hearer cannot contradict each other, i.e., $\exists v \in \Omega : P_S{}^\sigma(v) > 0 \wedge P_H{}^\sigma(v) > 0$. For a full theory, we need a general method for finding Nash equilibria (S, H) of speaker and hearer strategies which provide a *solution* to a given set of support problems $\mathcal{S}$. As argued for in Benz & van Rooij (2007), we assume that backward induction is used, which replaces the maxim of quantity. This means that the hearer will choose his actions from the set $\mathcal{B}(F)$ of all actions a for which $EU_H(a|[\![F]\!]) = \sum_v P_H(v|[\![F]\!])\, u(v,a)$ is maximal. The resulting hearer strategy in σ is a probability distribution $H^\sigma(\,.\,|F)$ over $\mathcal{B}(F)$. If the speaker knows the state of the hearer, i.e. if he knows σ, then he will optimise $EU_S(F) = \sum_v P_S(v) \sum_a H(a|F)\, u(v,a)$.

In our model of example (2a), we find that the speaker is uncertain about the hearer state. Let $[\sigma]_S$ be the subset of support problems which differ from σ only with respect to P_H. Then we can represent the speaker's uncertainty by a probability distribution $\mu^\sigma(\,.\,|v)$ over $[\sigma]_S$. It may depend on the state of the world v. In this case, the speaker will optimise $EU_S{}^\sigma(F) = \sum_v P_S(v) \sum_{\sigma'} \mu^\sigma(\sigma'|v) \sum_a H^{\sigma'}(a|F)\, u(v,a)$. The optimal answers Op_σ are then the answers $F \in Adm$ for which $EU_S{}^\sigma(F)$ becomes maximal. This has the consequence that $\mathrm{Op}_\sigma = \mathrm{Op}_{\sigma'}$ for all $\sigma' \in [\sigma]_S$. In e.g. Op_{+-}, $+-$ denoted the equivalence class of all σ for which $P_S{}^\sigma(S(a) \wedge \neg P(a)) = 1$. In our model (3), we assumed that the hearer knows the properties of b; this implies that the hearer state is determined by v and that therefore $\mu^\sigma(\sigma'|v) = 1$ or $\mu^\sigma(\sigma'|v) = 0$.[1]

We can represent the maxim of manner by the assumption that the speaker chooses optimal answers of minimal complexity. This defines optimal speaker strategies $S(\,.\,|\sigma)$ as probability distributions over Op_σ. It follows that an utterance of a form F carries the information that $F \in \mathrm{Op}_\sigma$. Hence, in σ, the hearer can infer from an utterance of F that a proposition R holds if $P_S{}^{\sigma'}(R) = 1$ for all σ' such that $P_H{}^\sigma = P_H{}^{\sigma'}$ and $F \in \mathrm{Op}_\sigma$. In order to count as an *implicature,* this fact must also be *common ground.* Let $O(F) = \{\sigma \mid F \in \mathrm{Op}_\sigma\}$ be the set of all support problems σ for which F is optimal. We denote the common ground updated with the information that F is optimal by $CG_\sigma(F)$ and define it as follows: Let $[\sigma]_S$ be as before, and let $[\sigma]_H$ be the analogue set of all support problems which only differ from σ with respect to P_S; let $E^0{}_\sigma := \{\sigma\} \cap O(F)$, $E^{n+1}{}_\sigma := O(F) \cap \bigcup\{[\sigma']_S, [\sigma']_H \mid \sigma' \in E^n{}_\sigma\}$, and $CG_\sigma(F) = \bigcup_n E^n{}_\sigma$. We tacitly assumed here that in all σ each interlocutor Y believes that all $\sigma' \in [\sigma]_Y$ are possible. In analogy to logical notation, we write $(\mathcal{S}, \sigma) \mid\!\models F \mathrel{+\!\!>} R$ if the *utterance* of F implicates that $\sigma \in R$ for sets of support problems $\mathcal{S}$ and $R \subseteq \mathcal{S}$. This leads to the following definition:

(4) $(\mathcal{S}, \sigma) \mid\!\models F \mathrel{+\!\!>} R :\Leftrightarrow CG_\sigma(F) \subseteq R.$

If we identify the meaning $[\![F']\!]$ of a form with the set $R_F = \{\sigma \mid P_S{}^\sigma([\![F']\!]) = 1\}$, then we arrive at the special case: $(\mathcal{S}, \sigma) \mid\!\models F \mathrel{+\!\!>} F' :\Leftrightarrow \forall \sigma' \in CG_\sigma(F)\, P_S{}^{\sigma'}([\![F']\!]) = 1$. If $(\mathcal{S}, \sigma) \mid\!\models F \mathrel{+\!\!>} R$ holds for all $\sigma \in \mathcal{S}$, then we can write $\mathcal{S} \mid\!\models F \mathrel{+\!\!>} R$.

In order to account for the suspension of implicatures, we assume that for each form F there is a set $M(F)$ which represents the normality assumptions about *utterances* of F. The set $M(F)$ is derived from a *preferential model* $\langle \mathcal{M}, \leq \rangle$ consisting of a partition $\mathcal{M}$ of a given set $\mathcal{S}$ of support problems and a linear well-founded order $\leq$. The preferential

[1] A more detailed model with uncertainty about hearer states was presented in Benz (2004).

model tells us which utterance situations are *normal* in general. For simplicity we assume that the speaker knows to which $M \in \mathcal{M}$ the actual support problem σ belongs, i.e., we assume $\sigma \in M \to [\sigma]_S \subseteq M$. We define $M(F)$ then as the $\leq$-minimal set M for which an utterance of F is *licensed.* The remaining problem is then to spell out what exactly *licensed* means. Before we address this problem, we adjust our definition of implicatures. If $M(F)$ is given, we can, in analogy to preferential nonmonotonic logics, characterise the preferred implicatures by:

(5) $\langle \mathcal{M}, \leq \rangle \mid\!\approx F +> R :\Leftrightarrow M(F) \models F +> R.$

Our final task is then to find a suitable definition of $M(F)$. In case of the model in (3), we can assume that $\mathcal{M}$ consists of the set M of support problems in which the speaker is an expert, and the set $\overline{M}$ of support problems in which the speaker is not an expert. The normal case is the expert case. From (5) it follows that it always holds that $F +> M(F)$. This entails that, if the speaker uses a form F which is optimal in a most normal situation, the addressee will infer that the situation is indeed most normal. Hence, if the speaker uses F in a non–normal situation, it will mislead the hearer if nothing stops him from thinking that the situation is normal. Let M_σ be the unique $M \in \mathcal{M}$ for which $\sigma \in M$. Then, the speaker's goal of avoiding misleading forms will lead him to restrict his choice to forms F which are not elements of any $\mathrm{Op}_{\sigma'}$ for which $M_{\sigma'} < M_\sigma$. This condition must be added to the definition of *admissible* answers Adm_σ. This addition entails that, in non–normal support problems, Op_σ will now contain other forms than before. This leads to the following definition of normal models *for utterances of F:*

(6) $M(F) = \min\{M \in \mathcal{M} \mid O(F) \cap M \neq \emptyset\}$

We finally show how this can be applied to the model in (3). We say that the speaker is an expert about a if $(\Box S(a) \vee \Box\neg S(a)) \wedge (\Box P(a) \vee \Box\neg P(a))$; hence, he is a non-expert iff $(\Diamond\neg S(a) \wedge \Diamond S(a)) \vee (\Diamond\neg P(a) \wedge \Diamond P(a))$. As $\Box\phi$ is equivalent to $P_S(\phi) = 1$, uttering, e.g., $\Box P(a)$ entails in conjunction with the assumption that the speaker is *not* an expert that $P_S(S(a)) < 1 \wedge P_S(\neg S(a)) < 1$, i.e., $\Diamond\neg S(a) \wedge \Diamond S(a)$. Hence, $\overline{M} \models \Box P(a) +> (\Diamond\neg S(a) \wedge \Diamond S(a))$. As we have seen before, $S(a)$, $P(a)$, $\neg S(a)$, and $\neg P(a)$ are all optimal answers for some support problem in which the speaker is an expert, but none of the boxed formulas $\Box S(a)$, $\Box P(a)$, $\Box\neg S(a)$, and $\Box\neg P(a)$ is one. Hence, the un-boxed formulas are not admissible in non-expert situations, but the boxed formulas are. As by assumption the hearer solves his decision problem by backward induction in which he only takes the literal meaning of a form into account, it follows that the more complex forms $\Box P(a) \wedge \Diamond S(a)$ or $\Box P(a) \wedge \Diamond\neg S(a)$ do not increase the speaker's expected utility. Hence, in all situations $\sigma \in \overline{M}$ in which the speaker knows that $P(a)$ but is not an expert, $\Box P(a)$ is an optimal answer, and it implicates that the speaker does not know whether $S(a)$. This explains the critical example (2a). We also find that the utterance $\Box\neg P(a)$ implicates that the speaker does not know whether $S(a)$, as do the answers $P(a) \wedge \Diamond S(a)$, $P(a) \wedge \Diamond\neg S(a)$, $\neg P(a) \wedge \Diamond S(a)$, and $\neg P(a) \wedge \Diamond\neg S(a)$.

REFERENCES

Benz, A. 2004. Questions, Plans, and the Utility of answers. In: C. Meier and M. Weisgerber (eds.). Proceedings of the Sinn und Bedeutung 8. 51–66.

Benz, A. and R. van Rooij. 2007. Optimal assertions and what they implicate: A uniform game theoretic approach. Topoi — an International Review of Philosophy 27: 63–78.

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

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

Hirschberg, J. 1991. A Theory of Scalar Implicature. New York: Garland Publishing.

Horn, L.R. 1989. A Natural History of Negation. Chicago: University of Chicago Press.

Levinson, S. 1983. Pragmatics. Cambridge: Cambridge University Press.

Levinson, S. 2000. Presumptive Meanings — The Theory of Generalized Conversational Implicatures. Cambridge, MA: MIT Press.

Schlechta, K. 2004. Coherent Systems. Amsterdam: Elsevier.

Shoham, Y. 1987. A Semantical Approach to Nonmonotonic Logics. In: K. L. Ginsberg (ed.). Readings in Nonmonotonic Reasoning. Los Altos, CA: Morgan Kaufmann Publishers, 227–250.