But What Do They Mean?
An Exploration Into the Range of Cross-Turn Expectations Denied by "But"

Kavita E. Thomas
School for Informatics, University of Edinburgh
kavitat@cogsci.ed.ac.uk

Abstract

In this paper we hypothesise that Denial of Expectation (DofE) across turns in dialogue signalled by "but" can involve a range of different expectations, i.e., not just causal expectations, as argued in the literature. We will argue for this hypothesis and outline a methodology to distinguish the relations these denied expectations convey. Finally we will demonstrate the practical utility of this hypothesis by showing how it can improve generation of appropriate responses to DofE and decrease the likelihood of misunderstandings based on incorrectly interpreting these underlying cross-speaker relations.

1 Introduction

In this paper, we will continue investigation into Denial of Expectation (DofE) across turns in dialogue when signalled by "but", following work by Thomas and Matheson (2003), and claim that these denied expectations need not be causal only. That is, we investigate two hypotheses: (1), that "but" can deny noncausal relations across turns in dialogue, e.g., temporal ordering relations, and (2) that because "but" is a negative polarity cue (Sanders et al., 1993), it inverts normal relations, and we will need to invert DofE dialogues in order to investigate the relations underlying the original (not denied) expectations.

To this end, we motivate the argument that these denied expectations can involve relations other than causal ones licensing the inference from A's turn to B's. We will then outline a novel methodology which utilises linguistic substitution tests on Knott's (1996) taxonomy of cue phrases to distinguish the underlying expectations denied. The practical utility of distinguishing these relations arises from discovering ways in which to both represent and utilise this information for NLG (among other applications), so we will address these issues in section 4. We show how the Information State (IS) (Matheson et al., 2000) representing the state of the dialogue in the PTT (Poesio and Traum, 1998) model of dialogue must be updated to reflect this new information, with Conversational Acts (Matheson et al., 2000) that do not simply indicate DofE as in (Thomas and Matheson, 2003), but also annotate the relation underlying the expectation being denied. Finally we will demonstrate how a system incorporating this information can improve generation of responses to DofE depending on its model of beliefs regarding the relation underlying the denied expectation in DofE dialogues.

2 Motivation

The main motivation behind modelling cross-turn relations is to get at what expectations and beliefs speakers might have upon interpreting the previous turn in the dialogue. Inferring the relations speakers perceive in cases where the related material spans speaker turns sheds light on how they interpret the previous speaker's turn, which in turn enables response generation that can specifically address these implicit relations. Here we focus on cases involving DofE, where the speaker of the "but" turn in dialogues like Ex.1 below has an expectation that *beautiful people* $>$ *marry*, where $>$ indicates defeasible implication.

(1) Example 1.
A: Greta Garbo was the yardstick of beauty.
B: But she never married.

Thomas and Matheson (2003) argue that B has the expectation that beautiful people (usually) marry, and interpreting A's utterance triggers this expectation, which B knows does not hold, since he knows that Greta never married, denying the consequent of the rule. Hence he generates DofE, and depending on A's beliefs w.r.t. B's assertion that Greta never married or the inferred expectation that beautiful people marry that is being denied, she can respond accordingly. E.g., if she agrees with the assertion but disagrees with the expectation, she can respond "But beautiful people don't have to marry!" Thomas and Matheson (2003) focus on modelling DofE

in Task-Oriented Dialogue (TOD). They present TOD examples like the following,

(2) Example 2.
A: Add the vinegar to the sauce.
B1: (Yeah) But it's not tangy enough.
B2: (Yeah) But we forgot to add the mushrooms.

where B1 involves an expectation similar to the one above involving beautiful people marrying, namely, that adding vinegar makes things tangy, which is a general cause-effect relationship. However they argue in that paper that B2 involves *satisfaction-precedence* (s.p.) between adding vinegar and adding mushrooms, namely, that B expects adding mushrooms to be done before adding vinegar. They then went on to argue that TOD DofE should be distinguished from Nontask-Oriented Dialogue (NTOD) DofE, because of examples like Ex.2B2 above, where the DofE arises from the denial of an ordering of actions in B's task-plan.

While we do not disagree with their claim that these s.p. DofEs in TOD (e.g., Ex.2B2) are distinct from causal cases like Ex.2B1, we disagree that these noncausal cases are unique to TOD; i.e., we argue for a unified treatment of DofE in TOD and NTOD, where, while search methods might differ (i.e., searching task-plans in TOD and private beliefs in NTOD), examples involving noncausal expectations which are denied are not unique to TOD. Consider the example below:

(3) Example 3.
A: Greta had a child in '43.
B: But she married in '47.

here we interpret B's "but" as signalling the denial of his expectation that marriage (usually) precedes having children in order to coherently interpret his response. The relation between turns (or antecedent and consequent) here is temporal ordering, and is very similar to the s.p. in the previous example (Ex.2B2). Unlike s.p., however, temporal ordering does not require the actions or states that temporally precede the later one to be achieved; i.e., the accomplishment aspect of s.p. is novel to planning, where goals are posted and accomplished, and there is a sense of agency. Temporal ordering relates actions, events, states, effects, etc, with no notion of agency involved. Prior work on DofE has not focussed much on the nature of the relation underlying the denied expectation, and we argue that this information will facilitate much more adaptive and appropriate response generation.

3 A Methodology for Distinguishing the Underlying Expectations

We outline a novel methodology for distinguishing features involved in these relations using linguistic substitution tests involving the cue phrase taxonomy presented in Knott's thesis (1996). Knott presents a taxonomy of cue phrases distinguished as feature-theoretic constructs rather than markers of one or more of a set of rhetorical relations as postulated in RST. Rather than finding data to describe a conceptualised theory of rhetorical relations, he uses data containing cue phrases to drive the creation of his taxonomy of cue phrases, which reveals psycholinguistic features involved in conveying or interpreting meaning, i.e., the data drives his theory of linguistic production. We enquire into the nature of the relations underlying these denied cross-turn expectations using the following methodology:

1. Take original "but" example and determine expectation being denied via algorithm in (Thomas and Matheson, 2003).

2. Invert example so that the consequent of the expectation is asserted rather than denied in B's turn, (i.e., omitting the "but"). (So the dialogue conveys that the expectation in Step 1 succeeds.)

3. Determine what sort of expectation this inverted pair of turns seems closest to, given Knott's taxonomy. Determine whether the cues conveying this relation are substitutable in this inverted dialogue:

   (a) Test all the high-level categories in the taxonomy and see which ones work by substitution tests involving cues belonging to those categories. Then determine whether the category chosen captures the nature of the expectation (i.e., intuitively, following annotator's judgment).

   (b) If so,

      i. test whether hyponyms[1] of these high-level cues work in the inverted dialogue. The most specific hyponyms that work indicate the maximally specific set of features that pertain to the relation underlying the expectation.

      ii. Now confirm that these cues that work in the inverted dialogue do not also work in the original (denied) dialogue. Those cues that work in the inverted example but not in the denied (original) dialogue are indicative of the nature of the underlying relation that's denied in DofE.

      iii. Look up the feature-value definitions for the maximally specific cues that work in this inverted (not denied) case. Comparing these to the feature-value definition of hyponyms of "but" that deny the same expectation will reveal which feature-values are denied/inverted in the denied case.
      Comparing the intersection of feature-values for hyponyms that work in the inverted case to the intersection of feature-values for hyponyms that work in denied case shows precisely which features are being denied.

   (c) If Knott's taxonomy does not provide a category that works for the inverted dialogue,

      i. check whether any of Knott's categories fit the original denied expectation by testing which cues are substitutable in the original example; a good place to start is with hyponyms of "but".

      ii. For cues that work, check their hyponyms to determine the maximally specific set of features that apply to the relation between turns. Note that this only specifies the relation underlying the denied expectation and does not shed light on the original (not denied) expectation.

      iii. If no cues besides "but" work in the original dialogue, then "but" must be the maximally specific cue that works, and we cannot determine more precisely the nature of the denied expectation, so assume that the turns are related by simple contingency/co-occurrence.

[1] Hyponyms inherit the features of their parent (higher-level) cues in the directed acyclic graph structure of the taxonomy. So all hypernyms (higher-level parents) should also be substitutable in the given case. Hypernyms are far less specific and therefore less precise.

Table 1: Feature-Values Denied in Ex.1

Features	Asserted *indeed* *even*	Denied *despite this* *then again*
Polarity	Positive	Negative
Source of Coherence	–	Pragmatic
Anchor	–	Cause-driven
Focus of Polarity	Anchor	Counterpart
Presuppositionality	Non-presupposed	Non-presupposed
Modal Status	Actual	Actual

3.1 An example

So to determine how A and B might be related (in B's perspective) for Ex.1, we find that the following cues work in the inverted dialogue below with the expectation asserted rather than denied:

(4) Example 4.
A: Greta was beautiful.
B: (Yes) <indeed> she <even> married.

The asserted expectation works with Knott's "additional information" category of cues, and "even" and "indeed" are the most specific of these cues which work. In the original denied example below, (with "*" indicating unacceptable cues):

(5) Example 5.
A: Greta was beautiful.
B: <However/even so/in spite of this/all the same/despite this/nevertheless/then again/*indeed> she never married.

two of the most specific of these negative polarity cues which work are "despite this" and "then again", which differ from "indeed" and "even" in polarity (the former are negative, the latter positive) and focus of polarity (the former are anchor-based, the latter, counterpart-based). Also, the cues which work in the inverted case do not work in the denied case. Furthermore, these negative polarity cues are defined for some values which are undefined for these additional information cues, and the two pairs also share some feature-values in common. But the features that are defined for both and differ are the ones which are the most informative; they specify which features are being denied in the DofE case, and which ones asserted in the inverted case. So here we find that DofE involves denying polarity and focus of polarity in the underlying expectation.

4 Modelling Issues

Although this methodology requires human judgment to assess the results of the substitution tests, it is a first step towards distinguishing underlying relations in DofE. We address how this information might be modelled in the PTT (Poesio and Traum, 1998) model of dialogue by adapting the Information State (IS) (Matheson et al., 2000) in order to facilitate more responsive generation from the system upon hearing the DofE.

4.1 Utilising Knott's Feature Definitions

Knott argues that his data-driven definition of relations is compatible with the view that relations are planning operators with preconditions and effects, where the relations' preconditions are defined via the speaker's intentions and applicability conditions specified for what the speaker wants to convey, and the effects are simply the intended effects the conveyed relation has on the hearer. More practically speaking, the features are defined in terms of variable bindings and relationships which describe the relations concisely. For example, the *polarity* feature describes whether the defeasible rule $P > Q$ holds, based on whether A=P and C=Q (positive) or A=P and C is inconsistent with Q (negative), where A and C are the propositional contents of the two respective related clauses. To address how polarity might be determined in a dialogue situation, if a speaker believes $P > Q$, then this is in her Private Beliefs field in the IS. If her turn is mapped onto Q, and the prior turn is mapped successfully onto P by matching first-order logic representations of the material in the two turns, then if her turn maps onto Q, we can assume positive polarity; if her turn maps onto a negated Q, then we assume negative polarity.

While mapping speakers' turns onto the variables which define Knott's features might be difficult, we can automate some of the feature assignment to update the IS by maintaining an exhaustive (i.e. complete) static table of cue-phrase definitions[2]. This way, once the most specific cue-phrases that work in the inverted and denied expectations are determined, we can automatically assign feature-value-pair bundles to these dialogues which describe the underlying relation being denied. Then comparing the feature-values for the maximally specific cues for both the asserted and denied cases (as we saw in the previous section), we can determine precisely which features are being denied in a given DofE, and the IS can be updated with this information, so that in the next turn of the dialogue, the speaker can compare these feature-value assignments to his own (in his private beliefs) and respond accordingly with a highly specific response to the DofE which targets precisely where he disagrees or agrees.

4.2 Information State Modification

We propose, given information about the nature of the underlying relation via the feature-value differences involved in the DofE as well as broader information about the category(ies) to which a cue-phrase belongs in Knott's taxonomy, to include this in the $dofe$ Conversational Act as follows for Ex.3:

[2]Knott provides a partial table of cue-phrase definitions like this in Appendix D.1 of his thesis.

$dofe(B, [temporal\_ordering(marrying(X,t), having\_children(X,t'); t < t')], [])$, where we replace $>$ with the more specific temporal ordering relation as the link between A and B's turns; the last field includes specific features being denied.

4.3 Responding to DofE Appropriately

Upon hearing B's DofE, A must then respond appropriately. If A also infers the nature of B's denied expectation, this can lead to much more responsive generation. (Thomas and Matheson, 2003) address how interpreting DofE in the IS can facilitate better generation. We argue that their algorithm cannot predict the correct expectations in cases involving noncontingency related expectations (i.e., cases unlike Ex.1). E.g., in Ex.3, their algorithm would predict that B has the expectation that "having a child in '43 $>$ not married in '47", since according to their original formulation of defeasible rules, B's turn is negated to form the consequent, so depending on A's beliefs he would respond accordingly. E.g., If A disagrees with both B's assertion and inferred expectation, then neither must be in his beliefs, and he might respond: "She didn't marry in '47, and anyway just because she had a child in '43 doesn't mean she should be married in '47." (I.e., A does not understand that B sees the events as temporally ordered.)

With our added information about the nature of this expectation, namely that it involves temporal ordering, we can improve upon Thomas and Matheson's scheme by predicting the following more appropriate responses. Notice that given this added information about the relation underlying the DofE, denying the DofE now means denying the underlying relation licensing the expectation. This means that A can be much more relevant when generating a response:

1. If A disagrees with both B's assertion and inferred expectation, then neither must be in his beliefs, and he might respond: "She married before '43, and anyway lots of people back then had children before marrying."

2. If A only agrees with B's assertion, then this assertion must be in his private beliefs, and he might respond: "Yeah, but lots of people back then had children before marrying."

3. If A only agrees with B's expectation, then this must be in his private beliefs, and he might say: "But she married before '43."

4. If A agrees with both B's assertion and expectation, he might say (minimally): "Yes, that's odd."

Notice that these responses indicate that A has understood B's temporal ordering that underlies these events and is the source of the denied expectation, and this allows B to correct possible misunderstandings. E.g., if B realises that A thinks she believes that people need to marry *before* having children, and this is an incorrect inference on A's part, she can indicate this by responding, e.g., to the situation in which A disagrees with both B's inferred expectation and assertion, B: "OK, but I don't think that people had to marry before having children." B needs to recognise specifically that A failed to interpret her temporal ordering expectation in order to correct A's misassumption. In cases in which specific features are the precise source of the DofE, if the hearer of the DofE can recognise that the wrong polarity is being attributed to his utterance, he (A) might indicate this misassumption by saying "but not marrying is common among beautiful people" (Ex.1).

5 Conclusions and Future Work

We present a novel treatment of DofE, in which we argue that the expectation denied in DofE across turns arises from a specific relationship between the antecedent and consequent. We then demonstrate a novel methodology for distinguishing the nature of this underlying relation via linguistic substitution tests on Knott's taxonomy of cue phrases. Finally, we show how this information can be used to generate more relevant responses that indicate explicitly what speakers have inferred from the preceding turn, allowing for faster detection and resolution of misunderstandings.

References

Alistair Knott. 1996. *A Data-Driven Methodology for Motivating a Set of Coherence Relations*. Department of Artificial Intelligence, University of Edinburgh.

Alistair Knott. 1999. *Discourse Relations as Descriptions of an Algorithm for Perception, Action and Theorem-proving. In Proceedings of the International Workshop on Levels of Representation in Discourse (LORID '99).*

Luuk Lagerwerf. 1998. *Causal Connectives Have Presuppositions*. Catholic University of Brabant, Holland Academic Graphics, The Hague, The Netherlands.

Colin Matheson, Massimo Poesio, and David Traum. 2000. *Modelling Grounding and Discourse Obligations Using Update Rules*. *Proceedings of the North American Association for Computational Linguistics.*

Massimo Poesio and David Traum. 1998. *Towards an Axiomatization of Dialogue Acts*. *Proceedings of Twente Workshop*.

T. Sanders, W. Spooren, and L. Noordman. 1993. *Towards a Taxonomy of Coherence Relations*. *Discourse Processes:15*.

Kavita Thomas and Colin Matheson. 2003. *Modelling Denial of Expectation in Dialogue: Issues in Interpretation and Generation*. Proceedings of the Sixth Annual CLUK Research Colloquium, Edinburgh, Scotland.

Kavita Thomas. 2003. *Modelling Contrast Across Speakers in Task-Oriented Dialogue: the Case of Denial of Expectation*. Proceedings of the 5th International Workshop on Multiple Approaches to Discourse (MAD'03), Driebergen, the Netherlands.