Modelling Correction Signalled by "But" in Dialogue

Kavita E. Thomas
School of Informatics
University of Edinburgh*
kavita.e.thomas@gmail.com

Abstract

Claiming that cross-speaker "but" can signal correction in dialogue, we start by describing the types of corrections "but" can communicate by focusing on the Speech Act (SA) communicated in the previous turn and address the ways in which "but" can correct what is communicated. We address whether "but" corrects the proposition, the direct SA or the discourse relation communicated in the previous turn. We will also briefly address other relations signalled by cross-turn "but". After presenting a typology of the situations "but" can correct, we will address how these corrections can be modelled in the Information State model of dialogue, motivating this work by showing how it can be used to potentially avoid misunderstandings. We wrap up by showing how the model presented here updates beliefs in the Information State representation of the dialogue and can be used to facilitate response deliberation.

1 Introduction

This paper addresses how cross-turn "but"[1] can signal correction in dialogue. We define correction as involving disagreement, denial or rejection of something in the previous turn and either an explanation for why this is disagreed with or the presentation of a replacement.

Although "but" shares the logical meaning of conjunction with "and", it carries a *conventional implicature* of contrast, as was argued by (Grice, 1975). (Carlson, 1985) argues that only in extreme circumstances does "but" introduce a flat denial, e.g., A says "he's dead," and B cries "but he's not dead!". He adds that it is relatively more common for "but" to contradict the preceding premise without restating the premise itself, as in Ex. 1 below.

(1) A: Nobody can do that.
B: But she did it.

He also claims that elliptic dialogue (see Ex. 2 below, B2) conveys an alternative rather than the direct contradiction which can be conveyed in B, and that B2 lacks emphatic force.

(2) A: He is extremely good.
B1: But he is slow.
B2: But slow.

(Kreutel and Matheson, 2001) show that "but" corrections can involve direct opposition by indicating contradictory evidence rather than directly negating a prior claim, and may be used to indicate disagreement:

(3) A: Helen didn't come to the party.
B: But I'm sure I saw her there.

We will start by describing the types of corrections "but" can communicate by focusing on the Speech Act (SA) communicated in the previous turn and address the ways in which "but" can correct what is communicated. We adhere to the idea that SAs act as wrappers around propositions (1962) and aim for a central definition of correction to arise from the analysis. To this end we will focus on previous turns communicating assertions, questions, commands and answers to questions and we will address whether "but" corrects the proposition, the direct SA or the discourse relation communicated in the previous turn. We will briefly discuss related relations signalled by cross-turn "but". After presenting a typology of the situations "but" can correct, we will address how these

*This research was conducted at the University of Edinburgh.

[1] We consider that "but" relates across turns if the constituent it modifies is contrasted with the previous turn. Turn-initial "but" almost always relates across turns.

brandial 2006: Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue, pages 34–41.

corrections can be modelled in the IS model of dialogue. We model correction in the Information State (IS) framework, assuming the PTT (Poesio and Traum, 1998) model of dialogue. We motivate our work by showing how it updates beliefs and facilitates response deliberation.

2 Distinguishing Different Types of Correction

Assuming that "but" corrects material in the previous turn, we will consider previous turns which are assertions, questions, commands and those which communicate implicit information like discourse relations. What we will investigate here is how cross-speaker "but" can correct utterances communicating different types of SAs. We investigate whether it corrects the proposition itself or whether it corrects the relations stated or inferred from interpreting the role of the SA.

2.1 Correcting Assertions

Assertions can involve a large range of propositions including propositions in which the clauses are related via local (intra-sentential) discourse relations e.g., causal relations, temporal ones, exemplification, etc. For example, speakers can assert that event X happened before event Y, or that something caused something else to happen, etc. Assertions can also be related to other turns in the preceding dialogue, or to information inferred from the context (including preceding discourse). For example, speakers can assert something as an example of some prior rule expressed in the dialogue, or as a cause for some prior event, etc. The example below involves A expressing a reason why chairs have four legs, which is a local causal relation.

(4) A: Chairs have four legs for stability.
B: But three-legged chairs can be equally stable, they just need to be well-designed.

Often asserted propositions relate to the preceding discourse, e.g., assertions that serve as answers to questions, as illustrated in the next subsection. So correcting assertions involves interpreting some relation asserted in the preceding turn, or the relation inferred by the corrector between the assertion and the preceding discourse or simply within the assertion itself (as above). Simply asserted material can also be corrected, as in the following two examples, where the assertions themselves (not relations within the assertion, as in the example above) are being corrected:

(5) A: Dogs are trustworthy animals.
B: (Oh,) But they just seem that way because they're dumb.
B′: No they're just too stupid to be devious.

(6) A: The train from Brisbane gets in at 11 o'clock.
B: (Oh,) But that's the Perth train! The Brisbane train only gets in at 1.
B′: No it doesn't, you're thinking of the Perth train.

Notice how both these examples are more definitively negated given a "no" as in B′. Indeed the "but" does not easily correct assertions since it logically communicates conjunction with the implicature of contrast/opposition. These "but" corrections of assertions often seem more understandable with the "Oh" beginning the turn, signalling surprise at the previous turn and indicating understanding and possibly also partial acceptance of the assertion, and disagreeing with either its content or role in the discourse. Disagreeing with an assertion's argumentative stance, usage, or role in the discourse (given discourse history) is similar to corrections of discourse relations which are discussed in the next few sections.

2.1.1 Correcting Answers

Answers to questions are SAs responding to preceding SAs (relational SAs, following (Poesio and Traum, 1998), and "but" can correct the answerhood relation itself, as presented in the correction conversation adapted from (Asher and Lascarides, 1998), which is simplified below:

(7) A: Why did (John get sent to jail)? r
B: (He was caught embezzling funds from the pension plan). p
C: Yes, (BUT (he went to jail) r because (he was convicted of tax evasion) $q1$) $q2$

The *why*-question here enables interpretation of B's turn as indicating a *reason* (and therefore an answer) to the situation for which A asks an explanation. We will take r to be the situation queried about in A, so r is "John got sent to jail". Interpreting B's assertion as an answer to the *why*-question, we have B communicating $reason(p, r)$. O Here the question sets up an expectation of an answer, which constrains interpretation. We interpret p as (1) an answer expressing (2) the reason for r and also, more basically (3), as an assertion of the situation described in B. Then C's "yes" is interpreted as accepting B's assertion, and the "but" clause ("but he went to jail because he was convicted of tax evasion", $q2$ in the example) indicates an alternative answer w.r.t. B's turn, since the first part of $q2$, "he went to jail" is r restated, and is explained (cued by "because") by $q1$ ("he was convicted of tax evasion"), which is expressed as a

Table 1: Graesser's Question-Types

Question	Abstract Specification
Comparison	How is X similar/different to/from Y?
Definition	What (category/properties) does X have?
Interpretation	How is an event interpreted/summarised
Feature Specification	What value/attribute does feature X have?
Causal Antecedents	What caused event X to occur?
Causal Consequents	What are the consequences of an event/state?
Goal Orientation	What are an agent's goals/motives?
Instrumental	How (plan) does an agent accomplish a goal?
Expectational	Why did some expected event not occur?

reason for r, so we get $reason(r, q1)$. Since we know that for a why-r question, anything which involves $reason(r, X)$ is an answer to the question, we can interpret this as an alternative answer to A's question. The "but" indicates that C's utterance is somehow contradictory to B's utterance, so we interpret this alternative answer as a correction of B, i.e., both (1) a rejection of B's utterance as an answer to A, and (2) the assertion of an alternative answer to A, which we will use as a rough definition for corrections like these.

The presence of the cue ("but") in Ex. 7 indicates that inferring that B's answer is wrong is obligatory[2]. However in either the cued or non-cued case, the inferred relation is still defeasible.

We take Graesser's taxonomy of inquiries (1992) as a basic set of question types and omit the categories in Graesser's taxonomy which involve single-valued (e.g., slot-filling) answers and prefer those which tend to require answers which attribute some predicate to a subject (so we have sentential rather than phrasal answers); see Table 1. Since several categories in his taxonomy have questions beginning with "what", some of which can also have *why*-questions, and the remaining three categories have *how*-questions, we will assume that we have the appropriate machinery to resolve several question-types. The benefit of using the taxonomy is that it provides us with useful clues about the nature of the answer, supposing the answerer to be honest and helpful (following Gricean reasoning).

2.1.2 Correcting Implicit Information

Implicitly communicated information like discourse relations, denied expectations, inferences, and defeasible rules can all also be corrected following our definition of correction where something is (1) either denied or rejected and (2) an alternative/replacement or explanation is presented. We will briefly explore correction of implicit information here.

Recall Ex. 4 where A's assertion communicates a reason for the stability of chairs, which is then refuted by B. Here B refutes the *reason* relation communicated by A by directly refuting the inference that A communicates that $four\_legs(chair) > stable(chair)$. Similarly, Ex. 7 showed that C could agree with the assertion B made but disagree with the inferred answerhood function of the assertion. The idea here is that implicit information like discourse relations can be corrected while not necessarily refuting what is explicitly asserted.

2.2 Correcting Presuppositions in Questions

What does it mean to correct a question? While there is much work involving the semantics of questions (e.g., (Asher and Lascarides, 1998), (Ginzburg, 1995), (Ginzburg, 1996), etc.), Asher and Lascarides argue that much of it falls either into the realm of dialogue planning or formal semantics, and neither type of approach bridges the gap in order to explain examples like the one below:

(8) A: How do I install the modem drivers in Linux?
B: But you're getting a Mac, so you don't need to install anything.

In order to get at how questions can be corrected, we will first consider some conditions on how they can be answered. Asher and Lascarides' approach to question-answering follows from the multiple notions of answerhood put forward in (Ginzburg, 1996), namely that (1) the information fully resolves the question, defined in terms of the interpreter's goal and mental state, or (2), that the information potentially resolves the question. That is, Ginzburg's notions of answerhood rely on context sensitivity and interpreter-specific responses, and his analysis identifies a proposition at the centre of the question, e.g., in the question above, the proposition would be "I install the modem drivers in Linux". In this sense, correcting a question itself would be very similar to correcting an assertion[3] except that it would involve adjustments to the answerer's obligations.

Additionally, different question types all expect specific answers, e.g., consider *why, how, what, when,* and *where* questions; in the case of "why X", the answer is usually a reason for X, where X is some proposition describing a state of events

[2]Thanks to a reviewer for this point.

[3]Thanks to Colin Matheson for this point.

or situation. "How X" expects an answer that provides a manner in which X might be accomplished, or a way to perform/achieve (X), where X is again a proposition containing a state of events. "What X" questions are less specific and X is often a phrase which is usually the subject of a transitive verb, so that the answer provides the direct object.

Of course these sketches of question-answering are very rough, and there is far more extensive work on the subject. The point here is that if we adopt Graesser's taxonomy of inquiries discussed earlier in the section on answers, we get much more specific requirements for answerhood, and also, a clearer set of question-types. In any case, our goal is to see how this information can inform a more specific characterisation of the types of corrections that are coherent given the preceding question's context. Given such specific information about what a question addresses and what sort of answer it expects, it becomes less difficult to see how the questions in Table 1 can be corrected. One way that correcting questions is different from answers is that the former involves pointing out how the question itself is invalid/incorrect/irrelevant (i.e., by providing an explanation for this incorrectness), rather than by disagreeing with and providing an alternative answer. Here are some example corrections responding to the first few categories in Table 1:

(9) Comparison: But X is the same as Y!
Definition: But X is undefined!
Interpretation: But it (the event) never happened!
Feature Spec: But I already told you!
Causal Ante: But nothing caused X to occur!

We notice here that corrections can deny the basis of the question, for example, that a comparison is valid in Comparison (above). Correction can also prove the question invalid, as in the Interpretation case, where it is impossible to interpret an event that never happened. Likewise, corrections can address meta-level issues as in Feature Specification above; here the corrector indicates that the question itself has already been answered. Notice the strong role played by the question category above; in many cases the correction hinges on the validity or relevance of the question category itself.

Another sort of correction of questions involves incorrect assumptions of slot-values in the question. Correcting misassumptions communicated in the question signals a difference in speakers' beliefs, as seen below:

(10) A: When did you want to fly back from Boston?
B: But I want to fly back from New York!
B′: But I want a rail ticket!
B″: But I don't want to leave Boston at all!

In B a slot-value is corrected, namely place of departure, and then presumably the question becomes valid. B′ indicates that the verb slot-value is incorrect; it should be "travel by train" instead. B″ questions the validity of the question itself by correcting the proposition at the centre of the question, since A asks when B wants to travel and B does not want to travel at all. This highlights an important point: questions presuppose the truth of their central propositions.

Precondition failure, constraints or mutually exclusive situations are also reasons why a question may be corrected. In the example below, B wants to go for a walk now and Hilda is not with B and therefore cannot come:

(11) B: I'm going for a walk.
A: Will you take Hilda with you?
B: But she's at school now and I can't wait.

B answers A while correcting A's assumptions of B's goals. We consider this correction because it involves rejection of A's inferred goal (namely that B takes Hilda with her) and also provides an explanation for why the question is deemed invalid. So the ways in which questions can be corrected are as follows:

- Correcting an incorrect slot-value (Ex. 10)
- Indicating that the question is not valid
  - Because a necessary criterion/precondition/ constraint has not been met (Ex. 11)
  - Because the question-type itself does not apply to what is being asked about (See 9, e.g.: Comparison, Definition, Causal Antecedent)
  - Because it was already asked and so is redundant (E.g., Feature Specification)

2.3 Correcting Commands

Corrections of commands also often involve correction of what is presupposed by the command. Interactions between speakers' plans often provide the source of disagreement evident in corrected commands. In the example below, B signals precondition failure (in order to shut the door, it must be open) and corrects A by asserting this.

(12) A: Shut the door.
B: But it's already shut.
B′: But then it'll get too hot; why don't we shut the window instead?
B″: But then it'll get too hot.

B′ communicates an undesirable effect of performing A's commanded action, and proposes an alternative. B″ just communicates the undesirable effect and does not propose an alternative solution, and we do not consider this correction. We will assume that correcting commands like questions and assertions also involves both (1) disagreeing, rejecting or denying something in the previous turn and (2) proposing an alternative/replacement or explanation. The question then arises as to why B above is considered a correction, since it does not provide an alternative. We argue that since B shows that A's action is invalid or impossible, it provides an explanation, similar to how the various question categories in Ex. 10 were deemed irrelevant or invalid. So in both these situations, the correction involves asserting what or

why the question or command is invalid or irrelevant. The difference between B″ and B above is that B″ only presents an undesirable effect and neither invalidates the command nor rejects and presents an alternative or explanation, so it is not deemed as correction. B on the other hand asserts that the commanded action is impossible (rejection/denial/disagreement) and presents an explanation, and so can be regarded as a correction.

Speakers can also propose better alternative actions, as seen below, where the rationale behind A's command is brought into question by the contradictory fact that the flies are getting in through the windows more than the door:

(13) A: Shut the door so the flies don't come in.
B: But more flies are coming through the windows; if we shut those instead we'll still have a breeze.

Speaker B can also object to the discourse relation inferred from A's assertion. In the example below, B infers that A's commanded action is intended to achieve the goal of keeping the flies out.

(14) B: These flies are really getting to me.
A: So shut the door.
B: But they are coming in through the windows.

In this case, what is actually being corrected is the inferred relation between the command and the problem mentioned in B(1). B(2) is correcting A's assumption that the flies are getting in through the door, and therefore corrects A's solution relation to B's problem, rather than the command A issues by showing that A's solution is irrelevant.

2.4 Related Relations

Correction differs from the other cross-turn "but" signalled relations concession and denial of expectation (DofE) in that it does not involve inferring relations between the turns themselves. Cross-turn DofE involves the "but" speaker denying an expectation triggered from the previous turn, while cross-turn concession involves the "but" speaker arguing in opposition to the other speaker w.r.t. a salient claim under discussion (Thomas, 2005).

Correction can also be distinguished from denial and rejection and seen as a relation which is composed of both denial or rejection and communication of an explanation or alternative. Many of the examples in the previous section involved rejection of offers, negotiation (by introducing alternative possible actions), misunderstandings, argumentation (by proposing other arguments), etc. We can distinguish denial as countering the perceived truth-value of an assertion, while rejection turns down an offer, and is thus only seen where commands or offers are issued, and both fail to introduce new explanatory or corrective information. For example, if C wants to deny B's answer in Ex. 7, he simply needs to say something along the lines of "No he wasn't (caught embezzling funds from the pension plan)". A rejection of a command like "shut the door" would simply involve an asserted "no!". Denial and rejection almost never involve "but", since they do not involve any acceptance beyond understanding of what is denied or rejected.

3 Modelling Correction

Given the discussion so far, we now show a generalised sketch of the update procedure for correction. We model correction in the Information State (IS) framework, assuming the PTT (Poesio and Traum, 1998) model of dialogue. In the procedure below, CDU, PDU and UDU refer to current, previous and ungrounded dialogue unit respectively. DH refers to the dialogue history field. CA_j etc. refer to conversational acts (CAs, see (Poesio and Traum, 1998)).

If [CDU.DH has CA_z of the form $assert(but[Z])$ AND a CA_y of [$reject(CA_j)$ OR $disagree(CA_j)$]] AND [PDU has CA_w of the form $SA(W)$]

1. If $alternatives(Z, W)$ then update CDU with CA_x: $correct\ (current\ speaker, W)$
2. Else if $explanation(Z, W)$ then update CDU with CA_x: $correct\ (current\ speaker, W)$

SA will be replaced by $command$, $question$, $assert$, or a discourse relation depending on what the SA of the prior turn (PDU) is. We assume a CA interpretation procedure (see (Thomas, 2005)) which takes CAs and compares their contents w.r.t. the speakers' beliefs, plans, etc. and updates the IS with $alternatives$ or $explanation$ depending on the case. Alternatives generally provide non-identical information relating to the same topic which is not necessarily conflicting or mutually exclusive, though in the former case, the corrector generally assumes that his alternative is more appropriate or relevant to the given situation. Explanations in the case of correction generally indicate why something is incorrect. In order to determine either alternation or explanation, the CA interpretator needs to call a theorem prover with the two related arguments. The above sketch of the update algorithm does not account for answers, which would need a question in UDU to be specified and where PDU and CDU pose as alternate answers.

Applying the update procedure to Ex. 7 produces the IS shown in Fig. 1, illustrating the results of the update procedure. We omit irrelevant fields and acts here for brevity, and just show

the relevant part of the IS[4]. The condition (COND field) says that if B accepts C's correction, then she is socially committed (see (Matheson et al., 2000)) to accepting C's answer.

We will now show a more specific procedure to address corrections of answers to why-questions:

If [UDU.DH contains an $ask(why[X])$ SA in CA_i] AND [PDU.DH has a CA_j that is of the form $assert(Y)$] AND [CDU.DH has a CA_z of the form $assert(but[Z])$] AND $CA\_interpreter\ (IS, CA_j, R)$ returns $R = answer\ (speaker(CA_j), CA_i)$ then

1. If CDU.DH contains a CA_k with the SA $assert(Z)$ and if the $CA\_interpreter\ (IS, CA_k, R2)$ returns $R2 = answer\ (speaker(CA_k), CA_i)$ and $CA\_interpreter\ (CA_j, CA_k, R3)$ returns $R3 = alternatives$, then add to CDU.DH CA_m: $reject(\ speaker(CA_k), R)$ and CA_n: $correct\ (speaker\ (CA_k), R)$. Add to CDU.COND $accept\ (speaker\ (CA_j), CA_n) \rightarrow scp(speaker(CA_j), R2)$

2. Else if CDU.DH contains a CA_k with the SA $assert(Z)$ and if the $CA\_interpreter\ (IS,\ CA_k, R2)$ returns $R2 = explanation\ (speaker\ (CA_k), CA_i)$, then add to CDU.DH CA_m: $reject\ (speaker(CA_k), R)$ and CA_n: $correct\ (speaker\ (CA_k), R)$. Add to CDU.COND $accept\ (speaker\ (CA_j), CA_n) \rightarrow scp(speaker\ (CA_j), R2)$

Notice that we do not address here whether the corrector accepts the assertion in PDU or not. We assume that the CA interpreter will determine disagreement when PDU is processed; i.e., it will check the speaker of CDU's beliefs, intentions, etc. to determine if there is any conflicts before updating the IS with accept/reject CAs. This procedure can be easily adapted to address questions of other types, and Graesser's taxonomy can be incorporated so that, provided we can parse questions into his categories, we know what sort of answers to expect, which can be utilised by the CA interpreter and theorem prover when determining if something is an answer or not. Lastly, we present the procedure for interpreting correction of questions:

If CDU.DH contains a turn-initial "but" and PDU.DH contains CA_j: $question\ (speaker[PDU], X, T)$ (where X is the proposition at the centre of the question and T is the question's category in Graesser's taxonomy)

1. If CDU.DH contains CA_k: $disagree\ (speaker\ [CDU], part\_of[X])$[5] and also CA_l: $assert\ (speaker\ [CDU], Y)$ and calling the theorem prover with $part\_of(X)$ and Y returns that they are alternate values for the same attribute or that both share a topic, then add to CDU.DH CA_m: $correct\ (speaker\ [CA_k], part\_of[X])$

2. Else if CDU.DH contains CA_k: $disagree\ (speaker\ [CDU], X)$ and also CA_l: $assert\ (speaker[CDU], Y)$ and calling the theorem prover with X and Y returns that they are alternate values for the same attribute or that both share a topic, then add to CDU.DH CA_m: $correct\ (speaker[CA_k], X)$

3. Else if CDU.DH contains CA_k: $disagree\ (speaker\ [CDU], CA_j)$ and also CA_l: $assert\ (speaker\ [CDU], Y)$ and calling the TA interpreter with TA_j (the TA equivalent for CA_j), X and Y returns that Y is a precondition/ constraint to be overcome of TA_j, then add to CDU.DH CA_m: $correct\ (speaker\ [CA_k], CA_j)$

4. Else if CDU.DH contains CA_k: $disagree\ (speaker\ [CDU], CA_j)$ and also CA_l: $assert\ (speaker\ [CDU], Y)$ and calling the theorem prover with CA_j and Y returns that $reason\ (incompatible\ [X, T], Y)$, then add to CDU.DH CA_m: $correct\ (speaker[CA_k], CA_j)$

Cases (a) and (b) above differ depending on whether X or part of X is disagreed with, as is illustrated in Ex. 10 turns B and B′ (part of X) and B″ (X) respectively, and the corrector provides alternate information (Y) for X with respect to the attribute they describe or the topic (T) that they relate to. Case (c) addresses situations in which the corrector introduces a precondition or constraint that must be satisfied before the question makes sense; for example, B could say "But I need to book when I'm leaving for Boston before I know when I want to leave" in response to A's question in Ex. 10. Case (d) addresses situations in which the question-type is somehow incompatible with what is being asked about (in the topic, T), and the corrector gives a reason about why this is the case, e.g., see Ex. 9 for an illustration of how different question types (following Graesser's categories given in Table 1) may be corrected. The point to using Graesser's categories here is to reduce the workload of the theorem-prover by checking for specific types of questions to isolate how they are corrected. On another point, it is important to note that none of these tests for correction will always be completely accurate, since contexts can always be found where such situations do not need correcting. However this is not something to worry about, because even if correction is erroneously predicted, the interpretation of such relations hinges on the notion that they are defeasible (i.e., cancellable), so that even if we have an incorrect interpretation, this can be cancelled explicitly by the correctee in a subsequent

[4]The abbreviated fields are: Previous and Current Dialogue Unit (PDU and CDU), Ground (GND) and Conditions (COND) following the IS structure given in (Matheson et al., 2000).

[5]Or $reject$; both are interchangeable for the purposes of this paper. The same holds for $alternate$ and $replacement$.

turn. In the next section we will see how the correctee can make use of the updates presented here to respond to the correction based on her own beliefs.

4 Deliberating Responses to Correction

The type of CA which is corrected in the $correct$ CA will give a lot of useful information about what the corrector believes so that the hearer can respond appropriately, depending on how her own beliefs relate to what she infers about the corrector. We will assume here that the most informative information available in the case of a correction will be found in the $correct$ CA itself, since this indicates which CA (in the previous turn) is being corrected, and further information as to why or providing an alternative will be found in an $assert$ CA along with the $correct$ CA. The focus on just three CAs greatly simplifies deliberation of a response to correction.

Now let us consider the various facets of response one can make to a $correct$ CA. Correctees can agree or disagree, they can do so partially or wholly, and they can agree or disagree with content explicitly or implicitly or with the relevance of the content. Furthermore, correction could involve new information to the correctee or highlight common knowledge and illustrate its relevance. In any case, the correctee must interpret what is being corrected and then respond, either by accepting that the corrector is right, or by disagreeing and explaining why.

Depending on what is corrected, the correctee needs to determine his own standpoint w.r.t. the correction by checking his beliefs, via the appropriate fields of the IS representing his beliefs, plans, what he has said, etc. Deciding to accept correction involves comparing the new (corrected) information with his original views in order to determine if the new information is compatible or not given his own beliefs and reasoning. While it is possible for a speaker to accept a correction and revise his beliefs, he should also be able to disagree if he knows of something wrong or incompatible in the corrected information. In practice this will involve theorem proving or reasoning about plans, so assuming such reasoning, we consider how one might deliberate a response based on the results of this reasoning. We will first discuss an example and then formulate a more general outline of what must be considered in responding to corrections. Recall Ex. 7. Here C accepted B's assertion p but disagreed with it as an answer to why John got sent to jail (r). An alternative reason for why he went to jail is also asserted by C (i.e., $q1$). If B accepts this reason, he can say so, e.g.: "Oh, I didn't know he also got caught for tax evasion." If B does not accept C's assertion ($q1$), he can refute it, e.g., by saying "No he wasn't. They dropped the charges." He can also accept C's reason and assertion but comment on the relevance of the correction and add new information, e.g., "Yes, but he also went to jail for embezzling funds from the pension plan and that's what A was asking about." We outline below some of the considerations a procedure modelling deliberation of responses to corrections should have:

If CDU.DH contains CA_j: $correct(speaker[CDU], CA_i)$, CA_k: $assert(speaker[CDU], X)$ and CA_l: $[disagree/reject(speaker[CDU], CA_i)$, and if PDU.DH contains CA_i:

1. Check PRIVATE BELIEFS, TASK BELIEFS (TB) and PDU.DH for speaker(PDU) for any CAs, beliefs, etc. held by speaker(PDU) which conflict with X. If there are conflicting beliefs update CDU with these conflicting beliefs. They should be added either to PRIVATE BELIEFS, TB or INT[6] (depending on where the conflict arose) as elements of the form $conflict(X, Z)$ where Z is the new information.

2. For all elements W of the form $conflict(X, Z)$ in CDU.PRIVATE_BELIEFS, CDU.TB or CDU.INT, push these onto INT in the form of intentions to $assert(speaker(PDU), W)$ (speaker(PDU) is the next speaker).

3. Push contents of PDU onto UDU and CDU onto PDU. CDU will contain the response to the correction, and will show the conflicting assertions in CDU.INT.

4. Deliberate over which intentions to assert should be expressed next (among other things) and generate a response.

5. After expressing $assert(speaker(CDU), W)$, remove $assert(speaker(CDU), W)$ from INT. Then move W into a CA of the form $raised[conflict(X, Z)]$ in CDU.DH.

This procedure first compares what is corrected and asserted by the corrector with the beliefs of the correctee in order to update the IS with a list of the resulting conflicting information found in DH (since the correctee might have previously uttered conflicting information), Private Beliefs in NTOD or Task Beliefs in TOD. The conflicts are then turned into intentions to assert them and pushed onto the correctee's INT field. After deliberating over which intentions to assert should actually be expressed next and these assertions of conflict are expressed, they are removed from INT

[6] If Intentions (INT field) are updated, they will trigger a response to be generated in the update rules.

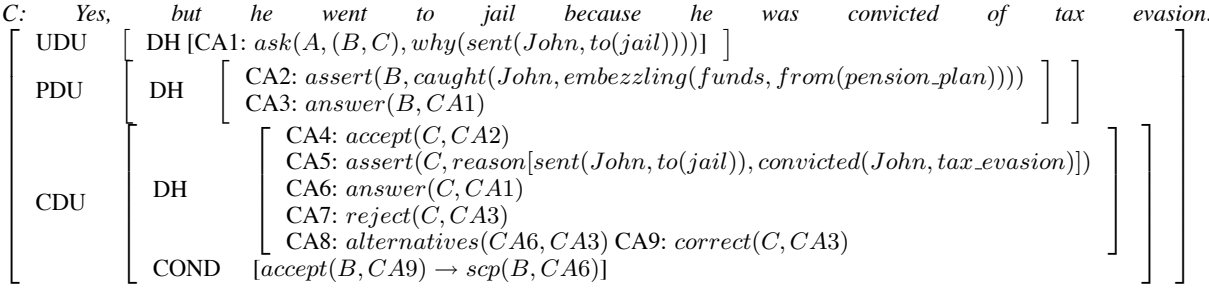

Figure 1: IS for Example 7

and CDU.DH is updated with a CA indicating that the conflict was raised. A benefit of this approach is that conflicts (raised by corrections in this case) are always straightened out without delay, hopefully reducing the number of misassumptions which might otherwise occur.

5 Conclusions and Future Directions

In this paper we present a novel treatment of cross-speaker correction when signalled by "but". We started out by presenting a range of corrections in assertions, answers to questions, implicit relations, questions and commands in the hopes that a central treatment for correction would emerge based on the notion that SAs are wrappers around a central proposition. We saw that both the SA and the central proposition can be corrected, and that in all cases, the defining characteristics of correction involve (1) disagreement with or rejection of something which is communicated in the previous turn and (2) the assertion of either an explanation or an alternative perceived to be more appropriate/relevant/correct by the corrector. We then presented procedures describing how correction can be interpreted given differences in beliefs etc. discernable by the theorem-prover and CA interpreter, and how this interpretation updates the IS representation of the dialogue with the correction relation. We finished by showing how these updates enable the correctee to respond to the correction depending on her own beliefs, etc. The motivation here is that by interpreting speakers' goals and beliefs and explicitly accounting for them, this approach helps to avoid potential misunderstandings.

In future work we hope to extend this analysis to account for multimodal corrections, in order to determine how the general approach can be extended to account for information communicated in nonlinguistic modalities as well as by linguistic means.

References

J. Allen and M. Core. 1996. DAMSL: Dialog Act Markup in Several Layers. University of Rochester.

N. Asher and A. Lascarides. 1998. Questions in Dialogue. *Linguistics and Philosophy vol.23:2*.

J. Austin. 1962. How To Do Things With Words. *Clarendon Press, Oxford*.

L. Carlson. 1985. Dialogue Games. *D.Reidel Publishing Company*.

J. Ginzburg. 1995. Resolving Questions, parts 1 and 2. Linguistics and Philosophy, vol.8.

J. Ginzburg. 1996. Interrogatives: Questions, Facts and Dialogue. In The Handbook of Contemporary Semantic Theory, Ed. S. Lappin.

A. Graesser, N. Person, and J. Huber. 1992. Mechanisms That Generate Questions. In Questions and Information Systems, Eds. T. Lauer, E. Peacock, and A. Graesser.

H.P. Grice. 1975. Speech Acts. *In Logic and Conversation, Eds. P. Cole and J. Morgan*.

J. Kreutel and C. Matheson. 2001. Incremental Information State Updates in an Obligation-Driven Dialogue Model. *Language and Computation*.

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

N. Lesh, C. Rich, and C. Sidner. 1999. Using Plan Recognition in Human-Computer Collaboration. *International Conference on User Modelling*.

D. Litman and J. Allen. 1987. A Plan Recognition Model for Subdialogues in Conversation. *Cognitive Science*, volume 11.

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

M. Poesio. and D. Traum. 1998. Towards an Axiomatization of Dialogue Acts. *Proceedings of Twente Workshop*.

E. Sweetser. 1990. From Etymology to Pragmatics. Metaphorical and Cultural Aspects of Semantic Structure. *Cambridge University Press, Cambridge*.

K.E. Thomas. 2005. *But What Do They Mean?* Modelling Contrast Between Speakers in Dialogue Signalled by "But". Doctoral thesis, *School of Informatics, University of Edinburgh, Scotland*.