Reconstructed Intentions in Collaborative Problem Solving Dialogues

Pamela W. Jordan
Intelligent Systems Program
University of Pittsburgh
Pittsburgh, PA 15260
jordan@isp.pitt.edu

Richmond H. Thomason
Intelligent Systems Program
University of Pittsburgh
Pittsburgh, PA 15260
thomason@isp.pitt.edu

Barbara Di Eugenio
Learning Research & Development Center
University of Pittsburgh
Pittsburgh, PA 15260
dieugeni@cs.pitt.edu

Johanna D. Moore
Computer Science Department
University of Pittsburgh
Pittsburgh, PA 15260
jmoore@cs.pitt.edu

Abstract

We provide evidence that speech act recognition, is 1) difficult for humans to do and 2) likely to misidentify proposals involving reconstructed intentions. We examine the reliability of coding for speech acts in collaborative dialogues and we present an approach for recognizing reconstructed proposals using domain context and other more easily recognized features.

1 Introduction

Speech act recognition plays a prominent role in dialogue understanding, in traditional approaches that infer a plan using plan construction operators [PA80], [LA90], [LC91, LC92], and in more recent techniques relying on statistical correlations or finite state machines [RM95, QDL+97]. Both approaches recognize surface speech acts, using surface form and information provided by the discourse context and the discourse operators, or by a finite state approximation of the planning information.

These approaches assume that it is (relatively) simple to recognize speech acts, and that speech acts are a required element of the analysis, corresponding closely to the speaker's intentions. In this paper, we provide evidence that:

(i) It can be very difficult for humans to reliably recognize speech acts, and

(ii) In some cases the association of a speech act with an utterance can reconstruct an intention far more determinate than anything the speaker entertained at the time of utterance.

In such cases, the problem solving context, together with certain surface cues that are not usually associated with speech act recognition, prove to be more predictive than the methods that are typically used in speech act recognition.

Cohen and Levesque [CL90] argue that it is unnecessary to recognize illocutionary force. Our conclusions agree with theirs in some respects, especially in the importance of the domain reasoning as evidence for interpretation. But, unlike them, we are not claiming that speech acts are a redundant and unnecessary level of analysis. And in fact, our examples seem to show that the agent intentions that underlie speech acts are not always based on simple, occurrent desires. Also, it is not clear how to fit joint speech acts into Cohen and Levesque's framework.

1.1 Joint speech acts

We have collected a collaborative problem solving corpus, in which subjects are asked to buy furniture for two rooms in a house. Later in the paper, we illustrate Point (i) by providing evidence from the coding of these dialogues.

Our tagging scheme is an augmented version of the Discourse Resource Initiative (DRI) one,[1] which is based on the familiar taxonomy of [Sea75].

Our coding difficulties derive in part from the fact that the usual taxonomies don't provide for *joint speech acts*. Some actions are joint, in that either they are literally performed by a group of agents rather than by a single agent, or they require a group's approval, even though a single agent may initiate the action [Tuo95]. An offer, for instance, is a speech act which attempts to secure a group commitment to some course of action. Without taking a stand on the status of joint actions, we treat them here as *sui generis,* without supposing they can be reduced to their component actions. Taxonomies like Searle's, which classify speech acts according to whether they impose constraints on the speaker S or the hearer H, do not fit joint speech acts or even their components, which in general impose coordinated constraints on both participants [Han79].

Amending Searle's taxonomy to provide for joint speech acts (as [Han79] suggests) helps to situate the issues. But this raises theoretical difficulties, by making the relation of speech acts to plans more complex and problematic. Planning formalisms are emerging (such as the SharedPlans approach),[2] and these could provide a theoretical framework for the recognition of joint as well as individual speech acts. Our experience with collaborative dialogues indicates that this would be a very welcome advance. But our evidence indicates that such a project may not solve the practical problems of recognition.

1.2 Reconstructed proposals

We will illustrate these problems with reference to RECONSTRUCTED proposals.

An agent engaged in collaborative problem solving must recognize when an agreement has been reached on sub-parts of the problem. One plausible approach to recognizing that an agreement has been reached is to recognize its components; propose and accept.[3] However this approach is not always reliable.

(i) A: I do not have a sofa for a better price but, i do have a

[1] See a manual draft at http://www.cs.rochester.edu:80/research/trains/annotation.

[2] [LGS90, GK93, GK95, Loc94, Loc95].

[3] Compare the approach of [Loc94, p. 28], which is essentially componential, in that the recognition of an intention to perform a (possibly joint) act depends on the prior recognition of a desire to perform that act.

lamp-floor, blue (250). i have a green table (200) and four chairs for (75) a piece.

(ii) B: ... the lamp and table sound good,

In sentence (i), agent A informs B that he has a floor lamp and in (ii) agent B takes this as a proposal to buy the lamp. A may well not intend to propose the floor lamp, but because B knows that she does not have a better alternative, she opportunistically treats (i) as a proposal and accepts it. If A doesn't object, the conversation continues as if A had indeed uttered an accepted proposal; and it is immaterial whether A definitely intended to propose the lamp when (i) was uttered.[4]

This is a matter of vague rather than of ambiguous intentions. So the interpretation can't just be a matter of choosing the most likely of several alternatives.

Our corpus makes it clear that reconstructed proposals are a natural and common way of doing business in some genres. Therefore, accounting for them is an important part of understanding the discourses in which they occur, despite the empirical problems of dealing with them.

1.3 Modeling the role of context

We believe that the problem of reconstructed proposals is connected with the the role of context in generation and interpretation. It has been known for a long while that speech acts are highly sensitive to features of the discourse context [Str64]. For instance, there are contexts in which an utterance like *I guess I can study tomorrow* is an acceptance of an offer. In these contexts, the main point of the utterance will be missed if this feature is ignored.[5] Even if speech acts are not recognized, context plays a fundamental role because it affects the effects a certain utterance has.

In our larger project, we are seeking to model interpretive reasoning of this sort using the following ingredients:

(i) A reasoning process that is somehow able to choose a likely and preferred interpretation from among several alternatives.

(ii) A model of the context in which the interpretation occurs.

[4] See [Fox87] for background on this sort of retrospective reinterpretation.

[5] Green and Carberry make a similar point about indirect answers in [GC94].

(iii) A mechanism for maintaining contexts, that allows them to be updated in the light of conversational information and common knowledge.

(iv) A general way of allowing these contexts to influence the interpretive process. In particular, contexts should influence the preferences that are assigned to interpretations.

Our approach[6] uses a form of weighted abduction as the reasoning mechanism [Sti88, HSAM93] with contexts modeled as modal-like operators. This allows contexts to be represented explicitly and reasoned about.[7] The abductive interpretation of an utterance will use a modal operator to limit the rules that are accessible and adjust the weights of assumptions. We can create contexts that favor the interpretation of an utterance like *I don't have any tables* as an acceptance. Similarly (since we also take an abductive approach to generation), we can also favor the generation of this utterance as a way of accepting in these contexts.

Our working hypothesis in modeling proposals is that the choice of context (ingredient iii) for this particular purpose is influenced, for both S and H, by the domain reasoning situation. In particular, if the suitable courses of action are highly limited, this in combination with coreference and surface form will make an utterance more likely to be treated as an acceptance (ingredient i); otherwise a propose must be made before an acceptance can be given. So the additional predictive power of the domain reasoning affects the context by adjusting the weights of assumptions (ingredient iv).

The remainder of the paper is organized as follows. In Section 2 we describe our domain; in Section 3 we discuss our evidence for the difficulty of coding for speech acts; in Sections 4 and 5 we return to the problem of reconstructed proposals.

2 The Collaborative Problem Solving Corpus

The subjects in our collected conversations are equal in status: they were both briefed on the domain knowledge needed for problem solving and neither is an expert at this task. The task is to buy furniture for the living and dining rooms of a house. (The task is based on those in [Wal93, WGR93]). Each subject is given a separate budget and inventory of furniture that lists

[6] See [TM95] for related discussion.

[7] See [MB95] for ideas about the formalization of context and contextual reasoning.

the quantities, colors, and prices for each available item. By sharing this information during their conversation, the subjects can combine their budgets and can select furniture from each other's inventories. The problem is collaborative in that all decisions have to be consensual; funds are shared and purchasing decisions are joint. Subjects are asked to maintain private graphical representations of their discussions and incremental agreements. We use this private information as partial evidence of what S's utterance meant and what H understood.

The subjects' main goal is to negotiate the purchases; the items of highest priority are a sofa for the living room and a table and four chairs for the dining room. The subjects also have specific secondary goals which further complicate the problem solving task. Subjects are instructed to try to meet as many of these goals as possible. The secondary goals are: 1) Match colors within a room, 2) Buy as much furniture as you can, 3) Spend all your money.

3 Identifying Speech Acts

3.1 Coding Schema

We devised our coding schema with two goals in mind: to conform as much as possible with the standards for mark-up being developed within DRI, and to represent the important features of the discourse that we collected. As with DRI, we tag aspects that are inherent to the utterance itself, and that encode the relationship of the utterance(s) to the preceding utterance(s). The tags of interest for this paper are:

(i) Utterance level tags: topic and illocutionary act.

(ii) Inter-utterance tags: relational tags (in particular, response-to) and coreference tags.

Topic tags[8] capture the meaning of the utterance by encoding what is relevant to the problem solving task in terms of furniture items or money: e.g., S may either state that he has a particular item (*I have a blue sofa for $300.*); discuss selecting a particular item (*shall we buy the two red chairs*); elaborate the description of an item that has already been introduced (*my red chairs are $100 each*); express an evaluation with respect to a specific

[8] We use the term *Topic* in a completely informal way.

furniture item (*the chairs seem expensive*); or discuss the budget (*I have $300.*).

Illocutionary-Act tags capture the intention behind the utterance and characterize at an abstract level S's main intention. At the highest level, the choices are *Inform, Directive, Commissive,* and *Conventional*.

An *Inform* utterance is intended to get H to believe something while a *Directive* is intended to get H to do something. Directives are further subdivided into: *Request-Action*, as in *Buy the chairs*; *Request-Info*, where the action requested is that H provides the desired information—many questions will fall under this category, e.g. *What do we have left if anything?*; and *Suggest*, which is a *Request-Action* that is conditional on agreement with H, as in *How about buying those two chairs.*

The primary purpose of a *Commissive* is to commit S (in varying degrees of strength) to some course of action. *Commissives* are subdivided into *Promise* and *Offer*. This distinction reflects the conditionality of S's commitment. S's commitment to an offer is conditional on H's agreement, so that the conversation will felicitously continue with H either accepting or rejecting. A *Promise* is not conditional in this way—or, if it depends on H's agreement, presupposes this agreement.

The distinction between *Directives* and *Commissives* is sometimes hard to draw. They are distinguished by S's degree of commitment to the action in question (under the assumption H will agree). With a *Directive*, S asks/orders H to perform an action, while a *Commissive* constrains S's own actions. As we pointed out in Section 1, joint speech acts complicate this picture. Rather than complicating the taxonomy by adding another category, we arbitrarily stipulated that proposals, such as *Let's buy the two chairs*, should be tagged as *Commissive*.

Relational Tags capture part of the relation between an utterance and the previous discourse: namely, an utterance or group of utterances $\{U_i\}$ can be unsolicited, or can respond to a previous utterance or segment.[9]

The two relational tags of interest are:

(i) *Initiate*—$\{U_i\}$ is unsolicited.

(ii) *Response-To*—$\{U_i\}$ is a response to a previous utterance or segment.

(ii.a.) *Answer*—answers a question, e.g. *No, the chairs are 100 each.*

[9] Space constraints prevent discussion of our treatment of segments.

(ii.b.) *Accept*—accepts the proposal or content of its antecedent, e.g. *The chairs sound good.*

(ii.c.) *Reject*—rejects the proposal or content of its antecedent.

Coreference Relations capture how furniture items discussed in one utterance are related to those previously discussed by means of the tags *sameItem, subset* and *mut(ually)-excl(usive). SameItem* is used when an utterance is related to its antecedent via the same item or set of items. For the current utterance to be tagged as *sameItem*, it must discuss exactly the same items as the antecedent, otherwise the tag *subset* is used. The tag *mut(ually)-excl(usive)* is used when U_1 mentions a set of items $\mathcal{S}_1$, U_2 provides an alternative $\mathcal{S}_2$ to that same set of items, and $\mathcal{S}_1$ and $\mathcal{S}_2$ are mutually exclusive.
Note that an *Initiate* utterance can still be linked to the preceding discourse via coreference relations.

3.2 Analysis of the Coding Results

We coded 5 of the 12 dialogues in our corpus (approximately 5,200 of 9,700 words). Two pairs of coders coded 2-3 dialogues per pair, with one dialogue in common to the two pairs. We report here on the pair-wise coder agreement, and also the agreement on the dialogue coded by three different coders. Table 1 reports values for the Kappa coefficient of agreement[Car96] for five different categories—the Kappa coefficient factors out chance agreement between coders. In each cell, the first number is Kappa, and the second number is the actual size of the coded data for that particular tag.

The first column refers to the highest level speech act coding, e.g. *Inform, Directive, Commissive* and *Conventional*; and the second column refers to the distinctions under *Directive* and *Commissive*. The other three columns refer to *Topic*, *Rel(ational)Tags*, and *Coref(erence) (Relations)*. The values for RelTags are computed by taking into account the different types of *Response-to*, i.e. *Answer, Accept, Reject.*

The discourse processing community is currently using Krippendorff's scale [Kri80] to assess Kappa's significance. The scale discounts any variable with $K < .67$, allows only tentative conclusions for K between .67 and .8, and allows real conclusions only for variables with $K > .8$. Thus, Table 1 strongly suggests that speech acts are an unreliable category. Admittedly, the low coding reliability may be due to either the inadequacy of the taxonomy, a

	Speech-Acts				
	Level 1	Level 2	Topic	RelTags	Coref
Pair 1 (LB)	.68 (127)	.59 (127)	.81 (126)	.74 (117)	.82 (112)
Pair 2 (LP)	.51 (64)	.45 (64)	.76 (62)	.77 (60)	.74 (58)
3-way coding	.62 (35)	.60 (35)	.79 (35)	.83 (33)	.88 (32)

Table 1: Kappa values

lack of clarity in the coding manual, or both. However, we can at least safely conclude that speech act recognition is a complicated enterprise, and thus in the next sections we will use the reliably coded features to show how they can be used to predict that an utterance is an acceptance.

4 The Problem Solving Model and Context

We will show here how the suitable courses of action (as described in Section 1.3) are classified as determinate or indeterminate.

The domain problem solving for the task described in Section 2 is more readily modeled as a constraint satisfaction problem than a planning problem since the temporal ordering of *buy* actions does not effect the solution. We view the problem space as a set of variables that must have a single value or a set of values of a certain cardinality assigned to them. Since the set of possible values is not known at the outset of problem solving, the model must recognize when to treat the set of values as open, when to treat it as closed and when to reopen it.

We use the SCREAMER constraint logic programming language [SM93] to model the problem solving. Although SCREAMER does not handle dynamic variables, we temporarily resolve this by setting up the variables and constraints anew with each utterance.

The input is limited to just the shared knowledge of S and H as an effect of the utterance, and comprises:

1. the variables being considered
2. the accumulated values for these variables
3. the current constraints

We provided the problem solving model with this input for each utterance and it gave the solution size for the unsolved variables as output. We

Relational Tag	Solution Size Determinate	
	Yes	No
Accept	11	1
Other	29	53

Table 2: Correlation of Acceptance and Solution Size

Relational Tag	Corefers to an utterance in a prior turn	
	Yes	No
Accept	15	2
Other	21	114

Table 3: Correlation of Acceptance and Coreference

characterize the output solution size as indeterminate if the result is 0 or the shared value set for some variable is open, otherwise it is determinate. For example, the solution is indeterminate if S supplies appropriate values for a variable but does not know what H has available for this variable (i.e. the value set is open).

5 Predicting Acceptances

To test our working hypothesis, we first look at the predictive power of each of the coded features that we expect to play a role and then at how a combination of these features increases the predictive power. We will look at solution size, coreference, and topic compared to acceptances.[10]

Table 2 shows first that there is a correlation between utterances tagged as acceptances and the solution size as an effect of the utterance ($\chi^2 = 13.57, p < .001, df = 1$). Furthermore it shows that an acceptance more frequently has a determinate solution size.

Also, we see that acceptances more frequently corefer to an item in a prior utterance (Table 3) and more frequently are about getting an item or evaluating an item (Table 4).[11] But alone, none of the features reliably predicts acceptances.

[10] Not all utterances have all four features coded.

[11] In these last two tables, some expected frequencies are too low to validly calculate χ^2. More instances of utterances are needed.

Relational Tag	Topic ∈ {getItem, eval or relate} Yes	No
Accept	12	0
Other	31	105

Table 4: Correlation of Acceptance and Topic

Relational Tag	Predictive Rule Applies Yes	No
Accept	10	2
Other	2	80

Table 5: Correlation of Acceptance and Predictive Rule

Finally, we combine these features into the following rule. An utterance is more likely to be an acceptance when it is determinate, is linked via coreference (sameItem, subset, mut-excl) to a prior turn and either:

1. the topic is about an evaluation of an item (eval, relate) or
2. the topic is about getting an item (getItem) and the utterance linked by coreference either has a getItem topic or is determinate.

As shown in Table 5, the above rule correctly predicts a majority of the utterances labeled as acceptances and falsely predicts 2 out of 82 other utterances as acceptances.

6 Conclusions and Future Work

We have argued that reconstructed proposals cannot be identified by speech act recognition and have shown that speech acts are more difficult to code than other features when joint actions are involved. We have presented a rule for predicting which utterances are acceptances based on domain context and the other more reliable features.

In future empirical work we plan to code the remaining corpus in order to further test the predictive rule. We will also consider how to code and process summaries, as a means of checking agreement. We are also implementing and testing a simulation of the domain reasoning and part

of the discourse generation and interpretation, including the planning and interpretation of proposals.

7 Acknowledgements

This material is based on work supported by the National Science Foundation under Grant No. IRI–9314961, "Integrated techniques for natural language generation and interpretation."

References

[Car96] Jean Carletta. Assessing agreement on classification tasks: the kappa statistic. *Computational Lingustics*, 22(2), 1996.

[CL90] Philip R. Cohen and Hector J. Levesque. Rational interaction as the basis for communication. In Philip Cohen, Jerry Morgan, and Martha Pollack, editors, *Intentions in Communication*, pages 221–255. MIT Press, Cambridge, Massachusetts, 1990.

[Fox87] B.A. Fox. Interactional reconstruction in real-time language processing. *Cognitive Science*, 11:365–388, 1987.

[GC94] Nancy L. Green and Sandra Carberry. Conversational implicatures in indirect replies. In Henry S. Thompson, editor, *Proceedings of the Thirtieth Annual Meeting of the Association for Computational Linguistics*, pages 64–71, San Francisco, 1994. Association for Computational Linguistics, Morgan Kaufmann.

[GK93] Barbara Grosz and Sarit Kraus. Collaborative plans for group activities. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 367–373, Los Altos, California, 1993. Morgan Kaufmann.

[GK95] Barbara Grosz and Sarit Kraus. Collaborative plans for complex group action. Technical Report TR–20–95, Center for Research in Computing Technology, Harvard University, Cambridge, Massachusetts, 1995.

[Han79] M. Hancher. The classification of co-operative illocutionary acts. *Language in Society*, 8(1):1–14, 1979.

[HSAM93] Jerry Hobbs, Mark Stickel, Douglas Appelt, and Paul Martin. Interpretation as abduction. *Artificial Intelligence*, 63(1–2):69–142, 1993.

[Kri80] Klaus Krippendorff. *Content Analysis: an Introduction to its Methodology*. Beverly Hills: Sage Publications, 1980.

[LA90] Diane Litman and James Allen. Discourse Processing and Commonsense Plans. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*. MIT Press, 1990.

[LC91] Lynn Lambert and Sandra Carberry. A Tripartite Plan-Based Model of Dialogue. In *ACL91, Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 47–54, 1991.

[LC92] Lynn Lambert and Sandra Carberry. Modeling Negotiation Subdialogues. In *ACL92, Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, pages 193–200, 1992.

[LGS90] Karen E. Lochbaum, Barbara Grosz, and Candace L. Sidner. Models of plans to support communication. In Thomas Dietterich and William Swartout, editors, *Proceedings of the Eighth National Conference on Artificial Intelligence*, pages 485–490, Menlo Park, CA, 1990. American Association for Artificial Intelligence, AAAI Press.

[Loc94] Karen E. Lochbaum. *Using Collaborative Plans to Model the Intentional Structure of Discourse*. Ph.d. dissertation, computer science department, Harvard University, Cambridge, MA, 1994.

[Loc95] Karen Lochbaum. The use of knowledge preconditions in language processing. In Chris Mellish, editor, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1260–1266, San Francisco, 1995. Morgan Kaufmann.

[MB95] John McCarthy and Saša Buvač. Formalizing context (expanded notes). Available from http://www-formal.stanford.edu/buvac., 1995.

[PA80] C. Raymond Perrault and James F. Allen. A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics*, 6:167—182, 1980.

[QDL+97] Yan Qu, Barbara Di Eugenio, Alon Lavie, Lori Levin, and Carolyn Penstein Rosé. Minimizing cumulative error in discourse context. In Elisabeth Maier, Marion Mast, and Susann Luper-Foy, editors, *Dialogue Processing in Spoken Language Systems*, Lecture Notes in Artificial Intelligence. Springer Verlag, 1997.

[RM95] Norbert Reithinger and Elisabeth Maier. Utilizing statistical dialogue act processing in Verbmobil. In *ACL95, Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, 1995.

[Sea75] John R. Searle. A taxonomy of illocutionary acts. In Keith Gunderson, editor, *Language, Mind, and Knowledge. Minnesota Studies in the Philosophy of Science, Vol. 7*, pages 344–369. University of Minnesota Press, Minneapolis, Minnesota, 1975.

[SM93] Jeffrey M. Siskind and David A. McAllester. Nondeterministic lisp as a substrate for constraint logic programming. In *Proceedings of AAAI-93*, CA, 1993. AAAI Press.

[Sti88] Mark Stickel. A prolog-like inference system for computing minimum-cost abductive explanations in natural-language interpretation. Technical Report 451, SRI International, 333 Ravenswood Ave., Menlo Park, California, 1988.

[Str64] Peter F. Strawson. Intention and convention in speech acts. *The Philosophical Review*, 59:439–460, 1964.

[TM95] Richmond Thomason and Johanna Moore. Discourse context. In Sasa Buvač, editor, *Formalizing Context*, pages 102–109, Menlo Park, California, 1995. American Association for Artificial Intelligence.

[Tuo95] Raimo Tuomela. *The Importance of Us*. Stanford University Press, 1995.

[Wal93] Marilyn A. Walker. *Informational Redundancy and Resource Bounds in Dialogue*. PhD thesis, University of Pennsylvania, 1993.

[WGR93] Steve Whittaker, Erik Geelhoed, and Elizabeth Robinson. Shared workspaces: How do they work and when are they useful? *IJMMS*, 39:813–842, 1993.