Tag-sets for the Generation of Nominals in Collaborative Dialogue

Pamela W. Jordan
Intelligent Systems Program and LRDC
University of Pittsburgh
pjordan@pitt.edu

Abstract

I describe the tag-sets I have used for testing algorithms and machine learning (ML) models that decide the content of nominals in collaborative computer-mediated dialogue. These tag-sets (Jordan, 2000b; Di Eugenio et al., 1998) that are based in part on the DRAMA (Passonneau, 1997) and DAMSL coding schemes (Allen and Core, 1997). Collectively the tag-sets provide information about the nominals expressed, the discourse relations between the nominals (e.g. coreference, set, class), the domain actions under discussion, and agreement states during collaborative negotiations as derived from influences on the speaker and hearer and the problem solving state as characterized by parameter settings. I end with an assessment of the usefulness of these tag-sets.

1 Introduction

In my work on generating nominals in dialogue (Jordan, 2000c; Jordan, 2000a; Jordan and Walker, 2000; Jordan and Walker, 2002), I've used manually-annotated, computer-mediated dialogues from the COCONUT corpus to compare the performance of various algorithms and machine learning (ML) models. The common input for all the algorithms and models is a representation of a discourse entity that is to be expressed at a particular point in a human-human dialogue. The common evaluation metric is a comparison of the attributes actually expressed in the human-human dialogue for the input discourse entity and the attributes that the algorithm or ML model selected for the same input entity. The input that differs for the algorithms and models is the other dialogue context that each deems necessary for making content selection decisions.

Clearly it will be difficult to address the question of what additional dialogue context is sufficient to annotate and make available for algorithms and models in general for an evaluative reference task (in the sense of (Whittaker et al., 2000)). But even the question of what annotations are necessary to arrive at the input discourse entities and output attributes that are common to all the algorithms and models is not without complications.

First I will provide some background on the COCONUT corpus and then I will describe the tag-sets that provide the input and output data for comparing performances. Next I will describe some tags for providing additional dialogue context. I will conclude with a summary of the performance differences measured using these tag sets and a brief analysis of the usefulness of the tag sets.

2 The Coconut corpus

The COCONUT corpus is a set of 24 computer-mediated dialogues consisting of a total of 1102 utterances. The dialogues were collected in an experiment where two human subjects collaborated on a simple design task, that of buying furniture for two rooms of a house (Di Eugenio et al.,

(In prior dialogue have already agreed upon: 3 green chairs for 50, 1 green chair for 100 and a green table in the dining room, and a blue sofa in the living room. The numbers in brackets indicate the utterance breaks for analysis of the dialogue.)

G: [36] That leaves us with 250 dollars. [37] I have *a yellow rug for 150 dollars*. [38] Do you have any other furniture left that matches for 100 dollars?"
S: [39] No, I have no furniture left that costs $100. [40] I guess you can buy *the yellow rug for $150*.
G: [41] Okay. [42] I'll buy *the rug for 150 dollars*. [43] I have *a green chair* [44] that I can buy for 100 dollars [45] that should leave us with no money.
S: [46] That sounds good. [47] Go ahead and buy *the yellow rug* and *the green chair*.
G: [48] I'll buy *the green 100 dollar chair*. [49] Design Complete?
S: [50] Sounds good, [51] do you want *the green chair* in the dining room with *the other chairs*? [52] I put *the yellow rug* in the living room. [53] Then the design is complete.
G: [54] Sounds good. [55] Hit the design complete

Figure 1: Excerpt of a COCONUT dialogue illustrating variable selection of attributes for nominal descriptions

2000). Their collaboration was via typed dialogue in a shared workspace where each action and utterance was automatically logged. An excerpt of a COCONUT dialogue is in Figure 1. A snapshot of the shared workspace for the COCONUT experiments is in Figure 2.

In the collected dialogues, the participants' 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 participants also have specific secondary goals which further constrain the problem solving task. The secondary goals are: 1) match colors within a room, 2) buy as much furniture as you can, 3) spend all your money. Participants are instructed to try to meet as many of these goals as possible, and are motivated to do so by rewards associated with satisfied goals.

Each participant is given a separate budget (as shown in the mid-bottom section of Figure 2) and an inventory of furniture (as shown in the upper-left section of Figure 2). Furniture types include sofas, chairs, rugs and lamps, and the possible colors are red, green, yellow or blue. Neither participant knows what is in the other's inventory or how much money the other has. By sharing information during the conversation, they can combine their budgets and select furniture from each other's inventories. Note that since a participant does not know what furniture his partner has available until told, there is a menu (see the mid-right section of Figure 2) that allows the participant to create furniture items based on his partner's description of the items available. The participants are equals and purchasing decisions are joint. Each set of dialogue participants solved one to three scenarios. These problem scenarios varied task complexity by ranging from tasks where items are inexpensive and the budget is relatively large, to tasks where the items are expensive and the budget relatively small.

3 Corpus Tag-sets

After the corpus was collected, it was first divided into utterance units and all the nominal expressions that were to be tagged were marked. Only the nominals that described task objects were marked. The definitions of utterance units and the nominal expressions to be marked are described in detail in (Jordan, 2000b).

Next the corpus was manually annotated with INPUT and OUTPUT tags as shown in Figure 3 and for additional dialogue context with DISTRACTOR SET and NON-IDENTIFICATION GOAL tags as illustrated in Figures 4 and 5. First, we will focus on the tags for creating the input common to all the algorithms and models that I tested. Next I will describe the tags used for comparing the output of each algorithm and model and computing the performance of each. Finally I will describe some of the additional dialogue context tags that were needed by the algorithms and models to make content selection decisions.

3.1 Input Tag-sets

The input for each algorithm or model is some representation of the object that is to be described at a particular point in the dialogue. A *discourse model* is used to keep track of the objects that have been discussed in a discourse. As an object is described, the conversants relate the information about the object in the utterance to the appropriate mental representation of that object in the discourse model (Karttunen, 1976; Web-

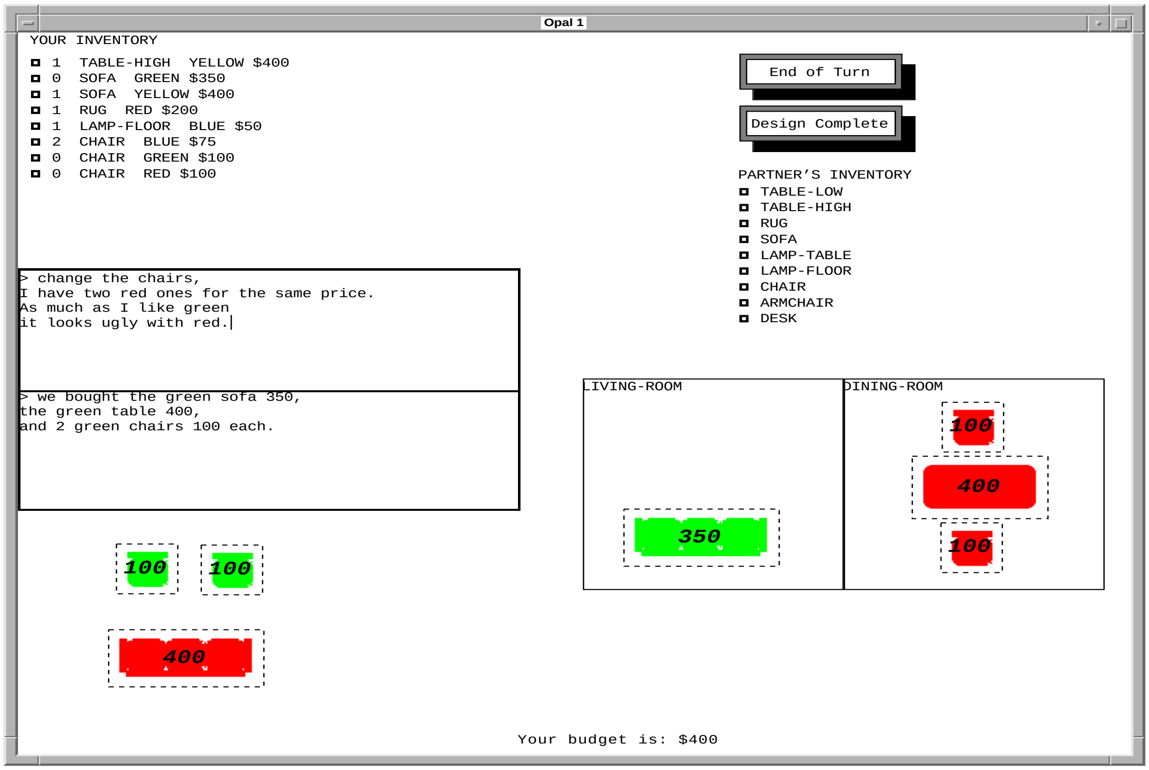

Figure 2: A snapshot of the interface for the COCONUT task

ber, 1978; Heim, 1983; Kamp, 1993; Passonneau, 1996). A *discourse entity* is a variable or placeholder that indexes the information about an object described in a particular linguistic description to the appropriate mental representation. We use the discourse entity as the common input representation.

Since a discourse entity is the product of a complex process of being engaged in a dialogue, our goal is to identify reliably recognizable tags from which a discourse entity can be automatically derived.

The INPUT tags in Figure 3 provide the discourse reference information which includes initial, coreference and discourse inference relations between different entities such as set/subset, class, common noun anaphora and predicative, and the attributes used to describe the entity. The discourse inference relations are based in part on the DRAMA coding scheme (Passonneau, 1997). For each tagged nominal in a dialogue, we either create a new discourse entity or update an existing one and link it to a copy of the unmodified entity. We update entities when a coreference relation is indicated by the tags and otherwise we create a new discourse entity.

When we create a new discourse entity, we start by adding its usage information to the representation; the utterance in which it occurred, how it was expressed and who used it in the dialogue. Next we add the tagged attribute information. We have three types of attribute tags for each entity:

- noun phrase level (e.g. *red* and *chair* in “the red chair”),
- utterance level (e.g. *I* in “I have a red chair”),
- locally inferred (e.g. *price* in “I can buy a blue chair with the money we have left.”)

To complete the discourse entity, we must next consider any relationships between the new entity and existing discourse entities. These relationships allow us to infer additional information about the entity we are creating.

To illustrate the discourse inference relations, in (1b), *the green set* is an example of a new discourse entity, which has a set/subset discourse inference relation to the three distinct discourse en-

Utterance	Reference and Coreference	Discourse Inference Relations	Attributes
37	initial ref-19	nil	my,1,yellow,rug,150
38	initial ref-20	nil	your,furniture,100
39	initial ref-21	class to ref-20	my,furniture,100
40	coref ref-19	nil	your,1,yellow,rug,150
42	coref ref-19	nil	my,1,rug,150
43	initial ref-22	nil	my,1,green,chair
44	corefers ref-22	CNAnaphora ref-22	my,100
47	corefers ref-19	nil	your,1,yellow,rug
47	corefers ref-22	nil	your,1,green,chair
48	corefers ref-22	nil	my,1,green,chair,100
51	corefers ref-22	nil	1,green,chair
51	initial ref-25	set of ref-12,ref-16	chair
52	corefers ref-19	nil	1,yellow rug

Figure 3: A dialogue excerpt and its input and output tags

tities for *2 $25 green chairs*, *2 $100 green chairs* and *$200 green table*.

(1) a. : I have [2 $25 green chairs] and [a $200 green table].

b. : I have [2 $100 green chairs]. Let's get [the green set].

A class inference relation exists when the referent of a discourse entity has a subsumption relationship with a previous discourse entity. For example, in (2) *the table* and *your green one* have a subsumption relationship.

(2) Let's decide on [the table] for the dining room. How about [your green one]?

A common noun anaphora inference relation occurs in the cases of one anaphora and null anaphora. For example, in (3) each of the marked NPs in the last part of the utterance has a null anaphora relation to the marked NP in the first part. Note that this example also has a class inference relation as well.

(3) I have [a variety of high tables] ,[green], [red] and [yellow] for 400, 300, and 200.

Discourse entities can also be related by predicative relationships such as *is*. For example, in (4) the entities defined by *my cheapest table* and *a blue one for $200* are not the same discourse entities but the information about one provides more information about the other. Note that this example also includes common noun anaphora and class inference relations.

(4) [My cheapest table] is [a blue one for $200].

For each of these relations, attribute values are inherited. The inheriting entity is the more specific of the two and attribute values are copied from the less specific entity when either the inheriting entity has no value for that attribute or has a less specific value (e.g. "superordinate" is a less specific value for the attribute *type* than "chair" is). The inheriting entity becomes the final discourse entity.

All but the predicative relation can involve more than two entities. When multiple entities are involved, we combine the entities and generalize values whenever attribute values differ (e.g. "red" and "blue" generalize to "range", "sofa" and "table" generalize to "superordinate" and "$50" and "$100" generalize to "$150" when the component entities do not have all the same attribute values). Inheritance then occurs between the combined entity and the entity being created. With common noun anaphora, only the *type* attribute is inherited. With the class relation, we further generalize the newly created discourse entity when it has been created from several component entities (i.e. we are creating the "class" entity in the instance/class relationship). To generalize a "class" entity, we removed generalized values and remove quantity values.

We only update an entity when a coreference relation is tagged. When we update the entity, we save a copy of the original entity and link it to the updated entity. By linking a copy of the original entity to the current one, we can trace how the entity evolved during the course of the dialogue and how and when it was used. Last, we update the usage information and make any changes in the attribute values that are indicated by the attribute tags.

For example, the initial representation for "I have a yellow rug. It costs $150." would include type, quantity, color and owner following the first utterance. Only the quantity attribute is inferred. After the second utterance the entity would be updated to include price because of the coreference and predicative relationships involved.

3.2 Output Tag-sets

The OUTPUT tag-set requires little interpretation to arrive at the output data we need for evaluating the performance of generation algorithms and models. We use the attribute tags in Figure 3 for both input and output determinations. As we mentioned earlier, there are three categories of attribute information that are tagged for each nominal; NP-level attributes, utterance-level attributes and locally inferred attributes. The NP-level attributes are always counted as attributes that a person chose to express to evoke an entity while the locally inferred attributes are not.

It is debatable whether the utterance-level attributes should count for description selection. Although we tagged whether or not the utterance-level attribute information is syntactically and semantically required, we have no way of capturing whether the requirements were met opportunistically or not. If we assume that the content for the nominal was selected first, then it is possible that syntax and semantics may have *moved* the attribute to the utterance level. For example, if the speaker chose ((owner A)(color red)(type chair)), then the syntactic and semantic requirements of "have" would make it unnecessary to also express the owner attribute value at the nominal level in "I have the red chair that we could use in the LR". However, if the speaker chose ((color red)(type chair)), the same sentence might still be produced to meet the syntactic and semantic requirements. Likewise the speaker might discard an attribute that was originally selected if the owner attribute value has the same power for uniquely identifying the target object.

To compare the performance of a generation algorithm to that of humans, we used a measure of the degree of match between the human's and the algorithm's selection of attributes for the same discourse entity in the same dialogue context. Inclusion and exclusion of an attribute both count in the degree of match. A perfect match means that the algorithm chose to include or exclude the same attributes as the human did for a particular entity. The measure, X/N, ranges between 0 and 1 inclusive, where X is the number of attribute inclusions and exclusions that agree with the human description and N is the number of attributes that could be expressed for an entity. We then did an analysis of variance and multiple pairwise comparisons on the match results to compare algorithms.

In the case of the ML models we required instead a perfect match between the human's and the ML model's choice of attributes. We then used k-fold cross validation to compare the performances of the different ML models.

3.3 Dialogue context Tag-sets

Most models for generating nominal expressions share basic assumptions about the speaker's goal when describing a discourse entity already introduced into the discourse model in prior conversation. The speaker's primary goal is *identification*, i.e. to generate a linguistic expression that will re-evoke the appropriate mental model. The description must be *adequate* for re-evoking the entity unambiguously, and it must do so in an *efficient* way (Dale and Reiter, 1995). One factor that has a major effect on the adequacy of a description is the fact that a discourse entity to be described must be distinguished from other discourse entities in the discourse model that are currently salient. These other discourse entities are called *distractors*. Characteristics of the discourse entities evoked by the dialogue such as recency and frequency of mention, relationship to the task goals, and position relative to the structure of the discourse are hypothesized as means of determining which entities are mutually salient

Utterance	Action	Introduce or continue	entity in Action
37	SelectOpt	intro act4	ref-19
38	SelectOpt	intro act5	ref-20
40	SelectOpt	cont act4	ref-19
42	SelectOpt	cont act4	ref-19
43	SelectOpt	cont act5	ref-22
44	SelectOpt	cont act5	ref-22
46	SelectOpt	cont act5	none
47	SelectOpt	cont act4	ref-19
	SelectOpt	cont act5	ref-22
48	SelectOpt	cont act5	ref-22
51	SelectOpt,	cont act5	ref-22
	SelectChairs	cont act3	ref-25
52	SelectOpt	cont act4	ref-19

Figure 4: A dialogue excerpt and its tags for deriving distractor sets

for both conversants. For this reason, we want to consider additional tags that provide information about the saliency of discourse entities.

Since a goal-directed view of sentence generation suggests that speakers can attempt to satisfy multiple goals with each utterance (Appelt, 1985) and since this also applies to lower-level forms within the utterance (Stone and Webber, 1998), it is reasonable to expect that a nominal expression can satisfy multiple goals. It is one possible explanation for why expressions that are non-minimal with respect to a goal to identify an object appear frequently in language. For this reason we also want to consider tags about the dialogue context that capture other communicative goals in addition to the identification goal.

3.3.1 Distractor Sets

In most computational work on generating nominal expressions, distractors are defined via a model of discourse structure. The most commonly used account of discourse structure for task-oriented dialogues is Grosz and Sidner's theory of the attentional and intentional structure of discourse (Grosz and Sidner, 1986). In this theory, a data structure called a focus space keeps track of the discourse entities that are salient in a particular context, and a stack of focus spaces is used to store the focus spaces for the discourse as a whole. The content of a focus space and operations on the stack of focus spaces is determined by the structure of the task. A change in task or topic indicates the start of a new discourse segment and a corresponding focus space. All of the discourse entities described in a discourse segment are classified as salient for the dialogue participants while the corresponding focus space is on the focus stack. These salient entities are treated as the distractors.

The DISTRACTOR SET tag-set in Figure 4 is used to capture the discourse structure. It encodes the goals and actions that are being discussed by the conversants and which discourse entities are arguments of these goals and actions. The tagged task goals are used to derive an intentional structure for the discourse, then a segmentation of the discourse, and finally the current focus space. The entity-action tagging helps indicate which entity belongs in which focus space.

But the definition of a focus space is still an open issue (Walker, 1996) since there is no clear criterion yet for assigning the segmentation structure. For this reason, algorithms and models may demand additional dialogue context in order to create alternative distractor set definitions. For example, two alternatives that we've used are based on extremely simple focus space definitions in which either the discourse entities from the most recent utterance or the last five utterances are possible distractors. Both of these alternative distractor set models are based solely on information about relative utterance distance.

3.3.2 Additional Communicative Goals

The tags for capturing non-identification goals are shown in Figure 5. They represent the problem solving state in terms of the influence the utterance has on both the listener and speaker, constraint changes that are implicitly assumed or explicitly stated by the conversants, and the size of the solution set for the current constraint equations, as well as current parameter assignments. The constraint change tags are goals that are used directly in the algorithms and models. The other tags are used to derive additional goals.

The solution set size for a constraint equation is characterized as being *determinate* if the set of values is closed and represents that the conversants have shared relevant values with one another. An *indeterminate* size means that the set of values in still open and so a solution cannot yet be determined.

Utterance	Influence on Listener	Influence on Speaker	Change in Constraints	Solution Size	Parameters
37	Action-Dir	Offer	drop color match	indet	OptionLR
38	nil	nil	color,price limit	indet	Option
40	ActionDir	Commit	none	det	OptionLR
42	ActionDir	Commit	none	det	OptionLR
43	OpenOption	nil	none	indet	OptionDR
44	ActionDir	Offer	none	det	OptionDR
46	ActionDir	Commit	none	det	OptionDR
47	ActionDir	Commit	none	det	OptionDR OptionLR
48	ActionDir	Commit	none	det	OptionDR
49	ActionDir	Offer	none	det	none
51	ActionDir	Offer	none	det	OptionDR
52	ActionDir	Commit	none	det	OptionLR

Figure 5: A dialogue excerpt and its tags for deriving non-identification goals

The influence the utterance is expected to have on the speaker and the listener, as defined by the DAMSL scheme (Allen and Core, 1997), helps capture some of the situational influences that may effect descriptions. The possible influences on listeners include *open options*, *action directives* and *information requests*. The possible influences on speakers are *offers* and *commits*. Open options are options that a speaker presents for the hearer's future actions, whereas with an action directive a speaker is trying to put a hearer under an obligation to act. There is no intent to put the hearer under obligation to act with an open option because the speaker may not have given the hearer enough information to act or the speaker may have clearly indicated that he does not endorse the action. Offers and commits are both needed to arrive at a joint commitment to a proposed action. With an offer the speaker is conditionally committing to the action whereas with a commit the speaker is unconditionally committing. With a commit, the hearer may have already conditionally committed to the action under discussion, or the speaker may not care if the hearer is also committed to the action he intends to do.

The influence on listener and speaker and the solution set size tags are used to derive the agreement state which encodes critical points of agreement during problem solving. According to (Di Eugenio et al., 2000), critical agreement states are:

- propose: the speaker *offers* the entity and this conditional commitment results in a *determinate solution size*.
- partner decidable option: the speaker *offers* the entity and this conditional commitment results in an *indeterminate solution size*.
- unconditional commit: the speaker *commits* to an entity.
- unendorsed option: when the *solution size is already determinate*, the speaker *offers* the entity but does not *commit* to using it.

From the agreement state for the current utterance and the previous agreement state we can derive additional communicative goals. As a first example, if a dialogue participant is *unconditionally committing* in response to a *proposal*, she may want to verify that she has the same item as her partner by repeating back the previous description.

Another example of derivable goals is situations where multiple proposals are under consideration that may contrast on attributes related to goals. For COCONUT these are either color-matching goals or price related goals. These contrasting goals depend on the agreement states, in that it is necessary to recognize proposals and commitments in order to identify alternatives and track agreed upon solutions. To recognize those dialogue contexts in which contrast goals can arise, we use heuristics involving the agreement states and parameter assignments to estimate partial solutions.

When there is a color contrast goal, it means that the entity's color matches with the partial solution that has already been agreed upon and/or contrasts with the alternatives that have been proposed. In this situation, there may be grounds for

Input-Output Tags	Reference and Coreference	Discourse Inference Relations	Attributes		
	.863 (z=19, p<.01)	.819 (z=14, p<.01)	.861 (z=53, p<.01)		
Distractor Set Tags	Introduce Actions	Continue Actions	Entity in Action		
	.897 (z=8, p<.01)	.857 (z=27, p<.01)	.857 (z=16,p<.01)		
Non-identification Tags	Influence on Listener	Influence on Speaker	Change in Constraints	Solution Size	Parameters
	.72 (z=19, p<.01)	.72 (z=13, p<.01)	.881 (z=11, p<.01)	.8 (z=6, p<.01)	.74 (z=12, p<.01)

Table 1: Kappa values for the Tag-sets

endorsing this entity relative to the alternatives. For example, in response to S's utterance [37] in Figure 1, in a context where G earlier introduced one blue rug for $175, G could have said "Let's use my blue rug." in response. In this case the blue rug would have a color contrast goal associated with it because it has a different color than the alternative, and it matches the blue sofa that had already been selected.

A price contrast goal can arise in two different situations. It exists either when the entity has the best price relative to the alternatives, or when the problem is nearly complete and the entity is more expensive than the alternatives. In the first case, the grounds for endorsement are that the item is cheaper. In the second case, it may be that the item will spend out the remaining budget which will result in a higher score for the problem solution.

Additional derivable communicative goals are described in (Jordan, 2000c; Jordan, 2000a).

3.4 Tag-set Reliability

The tags all have good intercoder reliability as shown by the KAPPA values given in Table 1 (Di Eugenio et al., 2000; Jordan, 2000c; Krippendorff, 1980). These values are all statistically significant for the size of the labelled data set, as shown by the p values in the Table.

4 Comparing Generation Models

In (Jordan, 2000c; Jordan, 2000a), I used the input and output tags along with the distractor set and non-identification tags in comparisons of Dale & Reiter's INCREMENTAL model, my INTENTIONAL INFLUENCES model and a model called RANDOM INFLUENCES that uses the INCREMENTAL model as a base and randomly adds additional attributes. My prediction was that the INTENTIONAL INFLUENCES model would do as well as the INCREMENTAL model but better than the RANDOM INFLUENCES model. The results confirmed my prediction; the INTENTIONAL INFLUENCES model was significantly better than the RANDOM INFLUENCES model and had a trend towards better performance compared to the INCREMENTAL model. Surprisingly, the INCREMENTAL model was not statistically different from the RANDOM INFLUENCES model.

In (Jordan and Walker, 2000; Jordan and Walker, 2002), we also used these same tag sets to compare the performance of ML models. In addition to the above tag sets we also used information that was directly derivable as a result of maintaining a history based on the input tag sets, utterance numbers, problem numbers and speaker information. This additional information included such things as the relative distances between utterances, between usages of discourse entities and between agreement states, frequencies of attributes expressed in reference chains and of previous mentions, and comparisons of speakers relative to an earlier utterance or discourse entity. This time we found that the INTENTIONAL INFLUENCES model performed significantly better than the INCREMENTAL MODEL and that both ML models performed better than their algorithmic counterparts. This indicates that the INPUT and OUTPUT tag-sets and the NON-IDENTIFICATION GOAL tag-set are useful.

An additional surprise was that varying the models for computing the distractor sets within both the algorithms and ML models made no significant difference in performance. This means that the DISTRACTOR SET tags may not be necessary since the alternative distractor set models depend only on simple recency and this information can be easily derived by maintaining a history based on the input tags and associated utterance numbers.

We also compared the performance of a number of other algorithms and ML models and the only additional dialogue context required by any of these was available as a result of maintaining a history of the inputs.

References

James Allen and Mark Core. 1997. Draft of DAMSL: Dialog act markup in several layers. Available from http://www.georgetown.edu/luperfoy/Discourse-Treebank/dri-home.html, under *Tools and resources*.

Douglas E. Appelt. 1985. Some pragmatic issues in the planning of definite and indefinite noun phrases. In *Proceedings of 23rd ACL*.

Robert Dale and Ehud Reiter. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science*, 19(2):233–263, Apr–June.

Barbara Di Eugenio, Pamela W. Jordan, and Liina Pylkkänen. 1998. The COCONUT project: Dialogue annotation manual. Technical Report ISP Technical Report 98-1, University of Pittsburgh, December. [On-line] Available: http://www.pitt.edu/˜coconut/research-papers.

Barbara Di Eugenio, Pamela W. Jordan, Richmond H. Thomason, and Johanna D. Moore. 2000. The agreement process: An empirical investigation of human-human computer-mediated collaborative dialogues. *International Journal of Human-Computer Studies*, 53(6):1017–1076, December.

Barbara Grosz and Candace Sidner. 1986. Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12:175–204.

Irene Heim. 1983. File change semantics and the theory of definiteness. In R. Bauerle, C. Schwarze, and A von Stechow, editors, *Meaning, Use, and the Interpretation of Language*. Walter de Gruyter, Berlin.

Pamela W. Jordan and Marilyn Walker. 2000. Learning attribute selections for non-pronominal expressions. In *Proceedings of Association for Computational Linguistics 2000*, October.

Pamela W. Jordan and Marilyn Walker. 2002. Learning to generate nominal expressions: Experiments with the coconut corpus. in preparation.

Pamela W. Jordan. 2000a. Can nominal expressions achieve multiple goals?: An empirical study. In *Proceedings of ACL 2000*, October.

Pamela W. Jordan. 2000b. Instructions for coding object attributes, intentions and contextual constraints. [On-line] Available: http:www.pitt.edu/˜pjordan/papers.

Pamela W. Jordan. 2000c. *Intentional Influences on Object Redescriptions in Dialogue: Evidence from an Empirical Study*. Ph.D. thesis, Intelligent Systems Program, University of Pittsburgh.

Hans Kamp. 1993. *From Discourse to Logic; Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer Academic Publishers, Dordrecht Holland.

Lauri Karttunen. 1976. Discourse referents. In J. McCawley, editor, *Syntax and Semantics*, volume 7. Academic Press.

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

Rebecca J. Passonneau. 1996. Using centering to relax Gricean informational constraints on discourse anaphoric noun phrases. *Language and Speech*, 39(2-3):229–264.

Rebecca J. Passonneau. 1997. Instructions for applying discourse reference annotation for multiple applications (DRAMA).

Matthew Stone and Bonnie Webber. 1998. Textual economy through close coupling of syntax and semantics. In *Proceedings of 1998 International Workshop on Natural Language Generation*, Niagra-on-the-Lake, Canada.

Marilyn A. Walker. 1996. Limited attention and discourse structure. *Computational Linguistics*, 22(2):255–264.

Bonnie Lynn Webber. 1978. *A Formal Approach to Discourse Anaphora*. Ph.D. thesis, Harvard University. Garland Press.

Steve Whittaker, Loren Terveen, and Bonnie A. Nardi. 2000. Let's stop pushing the envelope and start addressing it: a reference task agenda for HCI. *Human Computer Interaction*, 15:75–106.