The obligations and common ground structure of task oriented conversations

Luis A. Pineda, Varinia M. Estrada, Sergio R. Coria

Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS)
Universidad Nacional Autónoma de México (UNAM)
luis@leibniz.iimas.unam.mx

Abstract. In this paper a theory of dialogue structure for task oriented conversations, and its associated tagging scheme is presented. The theory introduces two linguistic structures supporting the dialogue which are called the obligations and common ground. The analysis of a transaction of a corpus in task oriented conversations is also presented. The empirical work from which this theory evolved is also briefly discussed.

1 Introduction

In this paper it is postulated that transactions in task oriented conversations or practical dialogues [2] are supported by two linguistic structures which we call the obligations and the common ground. These structures are 'built' by the speech or dialogue acts performed by the conversational participants, and a task oriented transaction is successfully concluded when the construction of these two structures comes to an end too.

The structure of obligations involves the specification of intentions through the realization of speech acts by one conversational participant, and the satisfaction of such intentions through linguistic acts, or perhaps through acts expressed in alternative modalities. For instance, an action directive stated by one conversational participant creates the obligation on the other to perform the specified act, provided that social and other contextual conditions hold; also, an offer establishes the obligation of the speaker to do something, provided the offer is accepted by the hearer. The structure of obligations is defined as the relation between the speech acts that state this kind of intentions and the speech acts that satisfy them, within conversational transactions. The structure of obligations is based on such a strong traditions and social conventions that is even satisfied in non-cooperative conversations [9].

The common ground structure, on the other hand, is defined as the relation between the speech acts through which conversational participants make sure that they share a common set of beliefs and intentions, and understand the utterances performed by their partners as intended [5]. In an idealized conversation, every speech act is understood as intended as soon as it is performed, and an implicit

common ground is held between participants along the whole of the conversation; however, in real conversations, the communication flow is commonly interrupted as a result of a failure to agree with, or to understand, a speech act; when a failure of this kind occurs, the common ground needs to be reestablished, or at least a note about the failure needs to be taken, before the conversation can proceed. The common ground can be broken in two main types of situations: due to a lack of agreement between the conversational participants, and due to an understanding problem. In the former case a speech act is listened well but the hearer fails to agree with all of its content, as it is the case in confirmation questions that put on hold, instead of accepting, a previous assertion. The latter case is exemplified by situations in which the message is not clear due to noise, or not determined enough due to its vague nature, and explicit speech acts are required to restore the common ground and make sure the participants are engaged in the conversation; typical speech acts of this kind are acknowledgments, back-channels, repetitions, completions, etc.

Speech acts need to be distinguished from the utterances that express them, and the same utterance may express more than one speech act, possibly in different conversational structures. For instance, an 'okay' may express a commit in the obligations structure, an accept in the agreement plane, and an acknowledgment in the understanding plane of the common ground.

There are constraints on the relation between speech acts; an action directive, for instances, needs to be paired with an action, and an information request with an answer; in the common ground, a hold act must be paired with an accept act, when the assertion that was put on hold is finally agreed upon, and an overt miss-understanding signal, like *what did you say*, must be paired an utterance that supplies the missing information. If a transaction satisfies all the stated constrains it is said that it is balanced, closed or completed.

Summarizing, we define the obligations and common ground structures as the relation between the speech in a conversational transaction, in addition to a number of constraints on such relation. In the rest of this paper, the specification of a theory of dialogue acts and conversation, based on the explicit realization of the obligations and common ground structure, is presented. In Section 2 the set of dialogue act types and the relations and constraints between them in the obligations and common ground structures is defined. In Section 3, the theory is illustrated with the analysis of a transaction of a task oriented conversation. In Section 4 a summary of the empirical work supporting the theory is presented. Finally, in Section 5, a general discussion of the theory and its potential applications for developing conversational systems is presented.

2　The transcription scheme

The notions of conversational obligations and grounding are explicit or implicit in a large number of theoretical studies (e.g. [5]); these notions have also been applied to the definition of dialogue managers [1, 9]; however, these structures are not reflected directly in annotation schemes, like DAMSL [6], which has been used for analysis of

dialogues with the purpose of specifying performance goals for conversational systems [1]. DAMSL distinguishes between the communicative status, the information level and the forward and backward looking functions of utterances, but discourse obligations and common ground acts are distributed implicitly in these four main dimensions. In particular, utterances expressing obligations, like action directives or information requests, are the prominent part of the forward looking functions, but there are also forward looking functions related to the common ground, like an affirm act introducing new information that eventually must be acknowledge by the hearer; conversely, although most explicit tags of the backward looking functions are mainly concerned with grounding, there are also some backward functions, like answers, that belong to the obligations structure.

In the present investigation we develop on DAMSL and, on the basis of the analysis of the DIME Corpus[1], the DIME-DAMSL tagging scheme has been introduced [8]. In this scheme, all four dimensions of DAMSL are considered, but in addition, the structure of obligations and common ground is made explicit; the specification of these structures includes the definition of an ontology of speech act types, and also the specification of the relations established through the realization of these acts. This relation is defined in terms of the 'charge' and 'credit' import of these acts; for instance, an action directive charges the obligation structures, and this charge is only 'credited' when the corresponding action is performed later on in the transaction. In accordance with basic accountability principles, a transaction is balanced when all the charges made in both the obligations and common ground structures have been credited. The current specification of the DIME-DAMSL scheme, developing on [8], is presented in tables 1 and 2.

Table 1. Balancing relations for the obligations plane

Charge	Time	Credit	On participant
Inf-request	I	Response	Other
Action-directive	I	Action	Other
Commit	I	Action	Same
Offer	P	Action	Same

Where:
- **Action** = {point-object | point-zone | point-path | point-coordinated-objects | place-new-object | move-object | remove-object | graph-plan | visual-interpretation}

In Table 1 it is also stated whether a charge is made at the time the speech act is realized (I) or whether the charge is postponed until the act is accepted by the interlocutor (P). The table also specifies whether the charge is on the hearer or whether it is made on the speaker himself.

The common ground is defined by agreement acts, related to the shared set of beliefs agreed along the dialogue, and by understanding acts, related to the

---

[1] http://leibniz.iimas.unam.mx/~luis/DIME/dimex-index.html

communication channel. In normal conversation, it is assumed that the content of an utterance is accepted by the interlocutor by default, and most forward looking obligation speech acts are accepted implicitly by the normal flow of conversation; however, there are also agreement acts that are expressed by explicit speech acts. We have observed three main cases: (1) the common ground has been broken, and needs to be repaired before the conversation is allowed to proceed, (2) the conversational participants are aware that some information is missing due to ambiguities or vague references, although the dialogue may be allowed to proceed, and (3) the common ground is reinforced by the explicit realization of speech acts.

Understanding dialogue acts express that the common ground needs to be strengthened; for instance, when the purpose of an utterance is to provide feedback from the part of the hearer to the speaker (e.g. acknowledgments, back-channels, etc.), reinforcing the belief of the speaker that the hearer is engaged; these acts make a communication charge that is implicitly credited by the interlocutor through the normal continuation of the dialogue. This level also includes explicit non-understanding signals that the common ground has been lost, like *what did you say*? The common ground relations are summarized next in Table 2.

Table 2. Balancing relations for the common ground plane

Charge	Credit	On participant
inf-request	Agr-action + affirm	other
action-directive	Agr-action	other
offer	Agr-action	other
open-option	Agr-action	other
affirm	Agr-action	other
reaffirm	Agr-action	other
vague-ref	fix	same/other
ambiguous-ref	resolve	same/other
Understanding-Act	Next (or current) utterance	other
Not-understanding-Signal (NUS)	Next utterance attending such signal	other

Where:
- **Agr-action** = { accept | accept-part | hold | maybe | reject | reject-part }
- **Understanding-act** = {acknowledgment | back-channel | repetition | rephrase | complementation | correction}

In the present formulation of this theory, and in addition to previous formulations [8], if the content of an obligation act is ambiguous or vague, this act makes a charge in the understanding plane too, and this act is credited when the ambiguity is resolved or the vague reference is fixed. Finally, in the common ground, a charge made by the speaker needs to be credited by the hearer, with the exception of ambiguities and vague references that may be resolved or fixed by the interlocutor that introduced them in the first place.

3   A tagging exercise

To illustrate this machinery, the analysis of a typical transaction of the DIME Corpus is presented in Table 3. In the figure, the numbers in the charge and credit columns index the utterance that expressed the corresponding speech acts, for the obligations, agreement and understanding planes respectively.

Table 3. Analysis of a transaction

U	T	Utterance (originally in Spanish)	Obligations		Common ground				Dialogue Act Types	
					AGR		UND		Obliga-tions	Common ground
			Ch	Cdt	Ch	Cdt	Ch	Cdt		
1	S	Do you want me to bring a piece of furniture to the kitchen?			1					offer
2	U	Yes	1			1			offer	accept
3		I need a stove	3		3				action-directive	action-directive
4	S	A second				3				accept
5		These are the five models of stoves that we have, simple stoves and stoves with lateral cupboards			5					open-option
6	U	Mmmm <sil> I'm going to select that stove			6	5				accept affirm
7	S	Okay				6				accept
8	U	eh, please I need it in <sil> in the far wall			8		8			affirm ambiguos
9	S	Which one is the far wall?	9						inf-req	hold
10	U	Let's see, here		9	10				answer	affirm point zone
11	S	There?	11						inf-req	repeat hold
12	U	Yes		11				8	answer	resolve
13	S	A second	13			10 8			commit	accept
14		<graphical action performed >		13 3 1	14				graph-action	affirm
15		Is there alright?	15						inf-req	
16	U	Yes, for the moment, yes		15		14			answer	accept

The first utterance in this transaction is an offer which creates a charge in the agreement plane, as offers need to be accepted or rejected; through utterance 2, *U* accepts the offer, crediting the agreement charge, and placing an obligations charge on *S*, as the system has now the obligation to perform the promised action; the main intention of the transaction is stated in 3 by *U*; this action directive places a charge on *S* in the obligations planes, and this charge is consistent with the offer made by *S* in the initial utterance. The action directive is explicitly accepted in 4, and this prompts the corresponding charge and credit (i.e. 3) in the agreement plane. Utterance 5 is an open option made by *S*; although this type of speech act is normally stated through a declarative statement, the utterance is not considered an affirm act,

as their purpose is not to enrich the set of beliefs of the interlocutor (i.e. to add a proposition in its knowledge base) but simply to allow him to choose from a predefined set of courses of action; also, the open option does not charges the obligation plane, as the interlocutor has no obligation to do anything about it; however, the open option does charge the common ground, as it needs to be accepted or ignored either explicitly or implicitly by the normal flow of the conversation, as it is the case in the present example. Next, *U* determines further the main intention through an affirm act in 6, and accepts implicitly the open option; although this act has, perhaps, an imperative connotation at the surface level, it is not considered an action directive (i.e. makes a choice); however, *U* needs to be sure that *S* took notice, and the affirm act charges the common ground; in 7, *S* accepts explicitly *U*'s choice, and credits the corresponding charge.

At this point of the transaction the main intention (i.e. to place a stove) and one of its arguments (i.e. what particular stove) have been fully determined, but the second argument, the location where the stove will be placed, is still to be specified. This is carried out from utterance 8 to 13. In 8 *U* states the desired location through an affirm act, with the corresponding charge in the agreement plane; however, the statement involves a spatial prepositional phrase (*on the far wall*) which is ambiguous; in the 2-D and 3-D views there are two walls that can be the referent, depending on the position adopted by the speaker in the virtual space; for this reason, 8 charges also the understanding plane with a ambiguous act.

In this situation, the conversation is allowed to continue despite of the ambiguity and, subsequently, the lack of a fully determined spatial referent; here, we hold that this is possible because the expression is understood in terms of its meaning, despite that its reference (the actual wall) is not singled out yet.

The spatial ambiguity is noticed by S and utters the confirmation question in 9, making the corresponding charge in the obligations plane, but in addition, it is also a hold act in the common ground that postpones accepting 8; in this situation the common ground is broken, and needs to be restored to continue with the dialogue; for this, a problem-solving process to resolve the referent of the remaining spatial argument is started. In 10, *U* answers 9 through an affirm act (i.e. *here*) at the time a spatial zone is pointed at (i.e. the zone corresponding to the far wall). The answer act credits the obligation plane, but the affirm act needs to be accepted and makes a new charge to the agreement level. The question in 11 expresses that the spatial reference needs still to be confirmed (*there*?), although *S* does not perform an explicit ostension, and puts on hold the affirm act in 10, in addition to its corresponding charge in the obligations plane. The answer in 12 credits this charge, and *U* resolves the spatial ambiguity that he himself had introduced in 8. Through 13, *S* accepts the postponed affirm acts in 8 and 10, which were uttered by *U*, making the corresponding credits to the understanding plane. At this point the main intention with its two arguments has been determined, and *S* is able to asses the task and to commit to do the action promised in 1 and accepted by *U* in 2. This concludes de intention specification phase of the transaction.

The satisfaction of the intention involves a problem-solving process that has the placing of the stove as its goal; this requires pairing the spatial referent introduced

with the pointing action in 10 with a reference position of the stove (e.g. the center or the bottom-left corner), and this involves the use of some design preferences and constraints adopted by the system. Finally, when the plan is decided, the actual action is performed and expressed through the graphical modality. This action credits the pending offer in 2, the action directive in 3 and the commit in 13 in the obligations plane. The graphical act makes also an affirm charge in the agreement level, as U needs to agree with the result of this action. To conclude the transaction, S makes a confirmation question in 15, creating the corresponding charge in the obligations plane, and this question is credited with U's answer in 16; finally, the graphical act is credited in the common ground with an accept act expressed by 16 too.

The transaction also shows that a commit act, at the boundary between the intention's specification and satisfaction phase, can be performed only when ambiguous referents have been resolved. In this case, the action satisfying the intention must be performed under certain amount of uncertainty, but its result must determine fully the spatial referent and the answer to the final confirmation question reestablishes the common ground.

This transaction does not involve vague expressions (e.g. *to the left of the stove*) where the spatial information needs to be further determined in order to undergo action, which are common in our corpus; however, these kind of expressions require fixing reference through a problem-solving process analogous the resolution of the ambiguity in the present example.

4   Tagging methodology and empirical work

The presented theory was developed in conjunction with a transcription exercise in which two dialogues of the DIME corpus were tagged and refined in several cycles by a team of taggers that at one point included 15 people. The exercise started from the original DAMSL scheme and its manual [6], and an initial course on this scheme was taken by all taggers. A dialogue was tagged by several people, and the kappa statistics was used to measure agreement between taggers [4]. The initial agreement scores were very low, specially for the common ground speech acts and the backwards dimension. One source of confusion was the implicit assumption that utterances express speech acts in a context independent fashion, as very few constraints between tags are defined in the original DAMSL scheme. In fact, the DAMSL manual provides explicit decision trees for agreement acts, and questions in these trees are focused on the function of a particular utterance, independently of its context in the dialogue. Also, as was mentioned, the obligations and common ground structures are only implicit in this scheme, and taggers easily confused the forward functions with the obligation structure, and the background functions with the common ground.

The theory presented in this paper evolved as reaction to these problems. Dialogues were first thought of as sequences of transaction; also, the obligations and common ground were made explicit, and the common ground was also explicitly

divided in the agreement and understanding planes of expression. Then, speech acts were classified according to these structures. In this exercise, the DAMSL dimensions (i.e. communicative status, information level and the forward and backward looking) were preserved, and the obligations and common ground structures were thought of as orthogonal to DAMSL dimension, enriching the level of structured postulated in the original scheme. The relations between speech acts within each plane of expression were modeled in terms of the charge and credit import of speech act types, and also in relation to the transaction context. In addition, as the DIME corpus is multimodal, tags for graphical actions and visual interpretations were included in the scheme.

In this exercise an Excel format was used to input the tags for al utterance in a dialogue; this format contemplated both the original DAMSL´s tags and dimensions as well as the obligations and common ground, and the charges and credits relations. The format also allowed the semi-automatic computation of the kappa statistics at three levels: transaction boundaries in the dialogue, charge and credits relations for the obligations, agreement and understanding planes of expression, and the actual speech act type tags. Through the exercise a number of conventions about the interpretation of speech acts in context, and also about the use of the tagging tools were defined and refined. The resulting scheme is called DIME-DAMSL.

With the tagging tool at hand, a formal experiment involving three tagging teams of three members each was developed. In this exercise two dialogues from the corpus were transcribed, in a sequence of tagging rounds; the teams were allowed to comment and discuss coincidences and discrepancies at the end of each tagging cycle and, after a few rounds, kappa statistics converged up 0.9 for transaction boundaries, charge/credits relations and the actual DIME-DAMSL tags. This figures suggests that the agreement between taggers above chance is very good, and that the tagging scheme and methodology are reliable. The version of the scheme presented in this paper includes, in addition, the understanding tags for charging vague and ambiguous references and crediting their resolution in the understanding plane. At the moment we have 10 dialogues tagged with the latest version of the scheme by two expert taggers, and although the strict kappa statistics has not been measured yet, between 84% and 94.5% of the utterances have received the same tags in transactions boundaries, obligations charges and credits, and also agreement and understanding charges and credits, in 5 dialogues (703 utterances) that have been compared. The current results support the case for the theory, and show that the tagging scheme is reliable. We plan to tag the 26 dialogues constituting the DIME Corpus in the near future, and compute the detailed kappa statistics for transaction boundaries, obligations and common ground balancing relations, and also for the actual speech acts expressed by each utterance.

5   Conclusion and applications

The analysis of speech acts is required in linguistic studies of discourse and conversation, and also for the construction of natural language conversational

systems, specially when spoken language is involved. In the present approach, the analysis of speech acts is partitioned in two levels: the level of form and the level of content. The level of form is constituted by the obligations and common structures, and this level is defined in terms of the relations and constraints between speech act in the context of the transaction, and these relations are independent of the actual conceptual content expressed by these acts. In this sense, the analysis of the obligations and common ground structure involves linguistic generalizations, that are independent of the particular propositions involved in the conversation, or knowledge about the application domain.

The main aim of the present view of dialogue structure is the construction of conversational systems in practical dialogues, where a dialogue can be analyzed as a sequence of task oriented transactions; we hold that typical transactions, in turn, can be modeled through dialogue models representing the obligations and common ground structures; in a finite state graph, for instance, states can represent conversational situations, and arcs the type of speech acts that relate situations. In this view, navigation through dialogue models depends on the ability to identify the speech act types expressed by utterance, taken advantage of the context and perhaps, of prosodic information [7]. We hold the hypothesis that this recognition is mainly a bottom-up process. We also hold the hypothesis that when dialogue act types are available, issues of content can be addressed in a top-down fashion; for instance, when it has been established that an utterances expresses an action directive in a given context, lexical and syntactic process can be directed to realize the specific action, as many other contextual aspects, like the agent and patient of the action, may be already available from the obligations structure; also, most common ground speech acts are interpreted within this level of structure, and this interpretation requires little lexical and syntactic processing.

In the present approach, issues related to discourse structure, reference resolution, both anaphoric or indexical, may also be simplified, as top-down interpretation processes focuses on the resolution of the arguments of specific instances of speech act types, when the type of the speech act in question is available already. In summary, the present theory is aimed to the construction of conversational systems in practical dialogues where the complexity of pragmatic inference can be reduced by the incorporation of dialogue models representing the obligations and common ground structures of typical transactions of the conversational domain.

In a more theoretical setting and according to the present view, a cooperative transaction can be seen as a cooperative problem-solving process in which an intention with its arguments is specified incrementally, followed by its satisfaction. In particular, expressions filling the intention's argument positions are initially understood through meaning, but such expressions have referents which need to be resolved, fixed or determined in order to act. The resolution of each of these arguments becomes embedded sub-problems, that are also solved cooperatively. The resolution of ambiguous or vague spatial referents, in particular, is an incremental process which is often concluded with an explicit ostension, and this deictic act restores the common ground and resolves reference in a single act. In this view, anaphoric and indexical resolution is subsumed in a process that aims to resolve the

referents of terms that are initially understood through their meanings, but lack a referent until this is resolved through a problem-solving process involving a context, and an interaction with the world.

Acknowledgments

References

1. Allen, J.F., Byron, D.K., Dzokovska, M., Ferguson, G., Galescu, L., Stent, A.: Toward Conversational Human-Computer Interaction. AI Magazine, 22(4):27–38, Winter, 2001.
2. Allen, J.F., Byron, D.K., Dzikovska, M., Ferguson, G., Galescu, L., Stent, A.: An Architecture for a generic dialogue shell. *Natural Language Engineering*, 6(3–4):213–228, 2000.
3. Allen, J.F., Core, M.G.: Draft of DAMSL: Dialog Act Markup in Several Layers Annotation Scheme. Department of Computer Science, Rochester University, October, 1997.
4. Carletta, J.: Assessing agreement on classification tasks: the kappa statistic. Computational Linguistics, 22(2):249–254, 1996.
5. Clark, H.H., Schaefer, E.F.: Contributing to Discourse. Cognitive Science, 13:259–294, 1989.
6. Core, M.G., Allen, J.F.: Coding Dialogs with the DAMSL Annotation Scheme. Department of Computer Science, Rochester University, 1997.
7. Coria, S., Pineda, L. Predicting dialogue acts from prosodic information. CICLing 2006, Lectures Notes in Computer Science, Springer-Verlag, Berlin Heidelberg, pp. 355 – 365, 2006.
8. Pineda, L. A., Castellanos, H., Coria, S., Estrada, V., López, F., López, I., Meza, I. Moreno, I., Pérez, P., Rodríguez, C. Balancing transactions in practical dialogues, CICLing 2006, Lectures Notes in Computer Science, Springer-Verlag, Berlin Heidelberg, pp. 331 – 342, 2006.
9. Traum, D., Allen, J. Discourse Obligations in Dialogue Processing. In Proceedings of Proceedings of *the 32nd Annual Meeting of the Association for Computational Linguistics* (ACL-94), pages 1–8, June 1994.