

Multi-agent human-machine dialogue: issues in dialogue management and referring expression semantics

Alistair Knott, Ian Bayard and Peter Vlugter

Dept of Computer Science, University of Otago

1 Introduction

Human-machine dialogue systems typically support dialogue between two agents: the human user is one agent, and the system plays the part of the other. In this scenario, the user and the system take turns at being the speaker, and when one of them is the speaker, the other is the addressee (the agent being spoken to).

However, in real life dialogue, there are frequently more than two participants. Automated dialogue systems can be configured in various ways to operate in a multi-speaker scenario. Firstly, a system can simulate each dialogue participant as a separate autonomous agent (e.g. Padilha and Carletta [1]). Secondly, a system can play the part of a single agent in a context where there are several human speakers (Wang, [2]). Finally, the system could support a dialogue between a single human user and several agents, all of which are played by the system. Here the agents can either be genuinely autonomous, or they can act in the service of a shared plan, delivering lines given to them by a central controller.

To extend a dialogue system to deal with multi-speaker interactions, whichever of the above scenarios is envisaged, a number of things must be supplied. At the dialogue level, we need a theory of turn-taking, to decide when to make an utterance, and who the addressees of other speakers' utterances are. At the level of sentence syntax and semantics, we need to pay special attention to constructions which are used to refer to dialogue participants (especially personal pronouns) and which are used to control turn-taking (especially terms of address).

We have already built a two-speaker dialogue system, which incorporates full sentence parsing and generation using a declarative grammar, and a range of standard dialogue management techniques (de Jager *et al* [3]; Bayard *et al*, [4]). This paper describes how we are extending this system to a multi-speaker environment, focussing on the additional syntactic constructions and dialogue management principles which are required, and on the interactions between these.

2 Te Kaitito: an English-Māori dialogue system

Our dialogue system, called Te Kaitito¹, supports bilingual human-machine dialogues in English and Māori. The user and the system alternate in generating contributions to a dialogue. When it is the user's turn to contribute, (s)he enters a sentence in English or Māori. The sentence is first parsed, using the LKB system (Copestake *et al*. [5]), and a set of syntactic analyses is computed. Each

¹ Online demos of Te Kaitito can be found at <http://tutoko.otago.ac.nz:8080/teKaitito/> .

analysis is associated with a semantic interpretation. One interpretation is then selected, using various principles (Knott and Vlugter, [6]). The dialogue manager then decides what kind of dialogue act is being performed by the utterance, and responds accordingly. If it encounters a problem interpreting the utterance, it responds with a suitable clarification question.

2.1 Presuppositional DRT

When an incoming utterance is parsed, its semantic representation is derived. The grammar associates sentences with representations in the Minimal Recursion Semantics (MRS) formalism (Copestake *et al.*, [7]). As a postprocessing stage, we convert these representations to a format called Discourse Representation Structures (DRSs; Kamp and Reyle [8]), with some additional annotations to deal with presuppositions and dialogue issues.

A DRS is a structure with two fields, one for representing **discourse referents**, and one for representing **conditions** or predications over these referents. DRSs are typically drawn as split boxes, where referents appear at the top, and conditions below. The discourse context is also represented by a DRS structure. This DRS represents the **common ground** of the conversation, in other words, the material which the speakers consider to be common knowledge between them.

A sentence's **presuppositions** are elements of its content which the speaker assumes are already part of the common ground. They are constraints on the kinds of context in which the sentence can be uttered. Here are two examples.

- (1) The dog chased a cat.
- (2) John's cat slept.

Sentence 1 presupposes that there is a dog in the discourse context (or more precisely, that there is exactly one *salient* dog in the context). Sentence 2 presupposes that there is someone called John, and also that this person has a cat. The DRSs for Examples 1 and 2 are shown in Figure 1. Notice that the presupposition DRSs are distinguished by dashed lines.

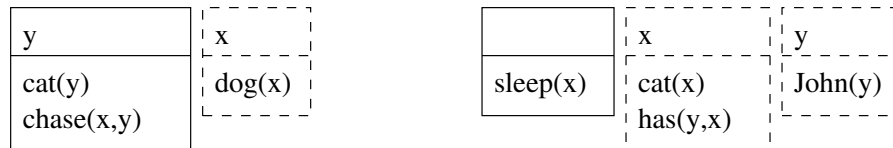


Fig. 1. *The dog chased a cat; John's cat slept*

2.2 Dialogue management concepts

Our dialogue system works with a number of different **dialogue acts**: assertions, questions, answers, acknowledgements and so on. We distinguish, as is traditional, between **forward-looking** and **backward-looking** dialogue acts.

A forward-looking act is basically a new initiative taken by a speaker; for example a new assertion made apropos of nothing, or a question about some new topic. A backward-looking act is one which is taken in response to an earlier dialogue act; for example the answer to a question, or the acknowledgement of an assertion. We will refer to a pairing of a forward-looking dialogue act and a backward-looking act as a **subdialogue**.

Another well-known idea in dialogue theory is that subdialogues can be nested inside one another. Here is an example of a subdialogue involving a clarification question.

- (3) Sid: can you pass the syringe?
Nancy: Which syringe?
Sid: The one on the left.
Nancy: Here you are.

The dialogue context can be thought of as containing a number of forward-looking acts which need to be responded to. These acts are sometimes referred to as the **dialogue stack**, or as a set of **questions under discussion**, unordered except for one **maximal** question, which is what a backward-looking act is assumed to respond to. In Example 3, after Sid's first utterance, there is one forward-looking act in the dialogue context, and after Nancy's first utterance, there are two, Nancy's being maximal. Sid's second utterance is therefore understood as responding to Nancy's question. In our system, we model the dialogue stack as a set of DRSs which are kept separate from the common ground.

3 Dialogue management in multi-agent dialogue

3.1 Deciding on the next speaker

Who the next speaker is allowed to be depends on what the previous dialogue act was. A simple rule is that if the previous utterance was a forward-looking dialogue act, the next speaker must be the addressee of that act. However, in nested subdialogues, this rule is not sufficiently general. For instance, in Example 3, Sid's second utterance is backward-looking; however, there is still a forward-looking act on the stack, which Nancy is bound to respond to. The general principle is as follows:

Principle 1: If the dialogue act which is maximal on the stack is forward-looking, the next speaker must be the addressee of that act.

When there is no forward-looking act on the stack—in other words, when the stack is empty—we assume anyone can talk next. Much has been written about how 'the floor is claimed' in such cases. In our system, we simply hand initiative to the user. The user can cede the floor, simply by pressing 'return' without entering a sentence. If this happens, the system currently selects one of the characters it plays at random to make a new initiative (either asking a new question, or making a new assertion).

3.2 Deciding on the addressee

The next issue is to determine the addressee of an utterance. If the speaker is played by the system, this issue is one of content selection, which relates to what the system wants to achieve. In our case, we keep things very simple, by always talking to the user.

Principle 2: If the system must make an utterance, the addressee is always the user.

If the speaker is the user, then determining the addressee is an interpretation issue. The remainder of this section sets out some principles which govern how this works.

Structurally-defined addressees In cases where there is a forward-looking dialogue act on the stack, we can use an analogue of Principle 1 to determine the addressee.

Principle 1a: If a speaker utters a backward-looking dialogue act, the addressee of this utterance is the speaker of the forward-looking dialogue act which is maximal on the stack.

In such a case, the addressee can be identified structurally, without being mentioned explicitly.

Explicit identifications of the addressee Other methods of indicating the addressee are more explicit. Firstly there are methods which are entirely non-verbal. For instance, when saying an utterance, the speaker can look at a particular person, or can even indicate one or more addressees by actually pointing at them. However, our system does not have the multimodal capabilities to simulate these nonverbal methods. Alternatively, the speaker can be even more explicit, and identify the addressee or addressees linguistically, using what we will call an **addressee term**.

Addressee terms can be used in two dialogue contexts. Firstly, if the addressee is already specified structurally, an addressee term can be given which is consistent with this.

- (4) Josephine [to Bert]: Shall we watch a video?
Bert: Not tonight, Josephine.² I have a headache.

If an addressee term is used which is inconsistent with the structurally specified addressee, we suggest the result is an incoherent dialogue.

- (5) Josephine [to Bert]: Shall we watch a video?
Bert: # Not tonight, Frank. I have a headache.³

² We believe that an explicit addressee term in such cases carries connotations of intimacy or of a heightened emotional connection. This seems a good example.

³ At best, Bert's reply can be understood on the premise that Josephine was asking on Frank's behalf. We do not consider this kind of proxy dialogue move; however, see Section 4 for a related construction we do cover.

The second context in which addressee terms can be used is at a point when the speaker is making a forward-looking dialogue act; in other words, where s/he is taking some new initiative.

Principle 3: If the speaker is making a forward-looking dialogue act, (s)he is free to choose any addressee or group of addressees.

For instance, in Example 6, Bob's first utterance is backward-looking, and has to be understood as being addressed to Sue even though there is no addressee term. But Bob's second utterance is forward-looking; he is thus free to address it to anyone, provided he makes this explicit with an addressee term.

- (6) Sue: Shall we go to the cinema tonight, Bob?
Bob: Good idea.
Bob: Do you want to come, Svetlana?

Note that Principle 3 as stated above should apply to forward-looking acts inside nested subdialogues; we expect that nested forward-looking acts should be addressable to any person. Indeed, nested forward-looking acts do seem to have this property, as the following example shows.

- (7) Sue: Shall we go to the cinema tonight, Bob?
Bob: Svetlana, do you want to come?
Svetlana: Good idea.
Bob: Okay.

Note that Svetlana's utterance closes her subdialogue with Bob, and thus that Bob's second utterance (a backward-looking act) has to be understood as being addressed to Sue, by Principle 1a.

Default addressees One final way of specifying an addressee is by default.

Principle 4: If a forward-looking act F is made without an explicit addressee term, then the addressee is taken to be the set of speakers involved in the previous subdialogue, minus the speaker of F .

Here is an example of Principle 4 in action:

- (8) Sue [addressing Bob and Mary]: Shall we go to the cinema tonight?
Bob: Good idea.
Mary: Good idea.
Bob: What film do you want to see?

Sue's first statement, together with Bob and Mary's responses to it, constitute a subdialogue. Bob's second utterance (*What film do you want to see?*) is a forward-looking act. Since there is no explicit addressee term, we assume by Principle 4 that it is addressed to Sue and Mary. Note that the principle also covers the case where the speaker was not involved in the preceding subdialogue. Here is an example of this:

- (9) Sue [addressing Bob]: Shall we go to the cinema tonight?
Bob: Good idea.
Mary: Hi there, what's up?

Mary in this case is interpreted as entering a dialogue whose participants are Sue and Bob. In this context, her utterance should be interpreted as addressed to both participants, unless she includes an explicit addressee modifier indicating otherwise.

4 Personal pronouns

In a multi-speaker context, there are some syntactic constructions which we need to pay special attention to, namely personal pronouns and addressee terms. We consider personal pronouns in this section, and addressee terms in Section 5.

Personal pronouns are devices which allow a speaker to refer anaphorically to him/herself, to the addressee, and to third parties. In each case, there is provision for the object of reference to be a single person, or a plural entity. The English personal pronoun system is quite simple, comprising first, second and third person pronouns, each of which can be singular or plural.

In other languages, the pronoun system encodes a a richer set of possibilities. For instance, in Māori, the language we are particularly interested in, there is a distinction between singular, dual and plural which is orthogonal to that between first and second person. Moreover, for first person plural, there is a further distinction between ‘inclusive’ and ‘exclusive’ pronouns, depending on whether the addressee is included in the set of people including the speaker.

Semantically, all pronouns introduce presuppositions about entities which are already in the discourse context. In the case of singular pronouns, the story is quite simple: first-person pronouns presuppose an object who is the speaker, and second-person pronouns presuppose an object who is the addressee. Consider the following dialogue:

- (10) Bob: I love you.
 Sue: I love you too.

Both of these sentences can be represented as the DRS in Figure 2 (left). Clearly,

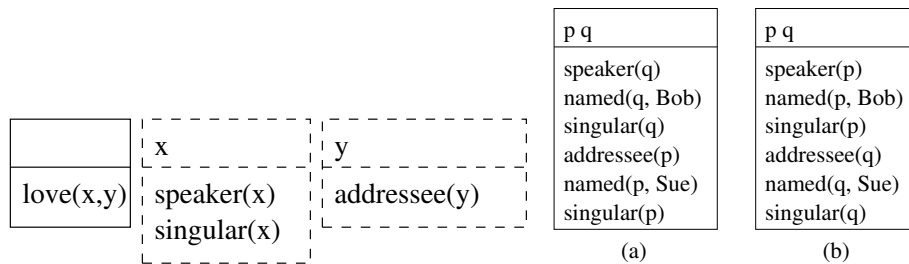


Fig. 2. DRS for *I love you*, with two possible contexts

in order to interpret such a DRS, the context needs to contain appropriate objects of which **addressee** and **speaker** are predicated. These predicates are unusual in two respects. Firstly, they need to change at each change of either speaker or addressee. For instance, when Bob’s utterance in Example 10 is interpreted, the

context DRS should be as in Figure 2(a), while when Sue’s utterance is processed, it should be as in Figure 2(b). Secondly, there can only be one **speaker** and one **addressee** predicate at any time; we assume there is only one conversation going on.

Plural personal pronouns To represent plural objects in general, we assume each plural object is associated with a discourse referent, whose members can be identified with a series of **member** predicates, and about which other special predicates such as **plural-object** and **cardinality** can be asserted. Plural personal pronouns denote—or more accurately presuppose—groups which include the addressee. For instance, *we* presupposes a group which includes the speaker, and plural *you* presupposes a group which includes the addressee. The group referred to by a plural pronoun must be made salient linguistically, for instance, by using a conjunction of noun phrases, as in Example 11:

- (11) Dean: Pearl and I went to the cinema yesterday.
Hank: What did y’all watch?
Dean: We watched ‘Casablanca’.

The first utterance here introduces a group entity composed of Pearl and Dean. In Hank’s responding utterance, *y’all* presupposes a salient plural entity one of whose members is the addressee (Dean); the group of Pearl and Dean satisfies this presupposition. In Dean’s second utterance, *we* presupposes this same entity.

This general definition of personal pronouns subsumes an interesting case where the speaker or the addressee are *themselves* group entities. Consider this case:

- (12) Bob: Sue and Mary, are you ready to go?

The group of people denoted by *you* in this example are not just being referred to, but being addressed; unlike the group denoted by *y’all* in Example 11, *you* in this example actually denotes the group of people who have to respond to the utterance. Note that the second person pronoun can still be defined as presupposing a salient group of people which includes the addressee, provided that (a) we are not talking about strict inclusion, and (b) we assume (reasonably) that the addressee of an utterance is always a salient entity in the discourse context.

According to our dialogue-management principles, a forward-looking utterance with a group addressee must be responded to by that group. Is it possible for an utterance to have a group speaker? Genuine ‘joint utterances’ are of course virtually nonexistent (outside the theatre). Our approach is to allow the possibility of group speakers, with a very simple additional dialogue management principle:

Principle 5: utterances made by a group speaker are actually made by an individual member of the group, on behalf of the whole group.

This approach is in fact in keeping with a general assumption in our system that all communication is public, and that there are no disagreements between participants. But in a more realistic situation, clearly much more needs to be said about how members of a group negotiate a response in such cases.

5 Addressee terms

Addressee terms function syntactically as sentence modifiers in English and Māori. In English, they can appear wherever a sentential adverb can appear; we believe they have the same distribution as conjunctive expressions like *however* or *therefore*. They are typically proper names (e.g. *Hello, John*) or bare nouns (e.g. *Hello, baby*). In Māori, they typically occur at the front of sentences:

(13) William, kei te auau ngā kuri. (William, the dogs are barking.)

One important exception: greetings are best modified postsententially:

(14) Kia ora William. (Hello William.)

What does an addressee term contribute semantically to a sentence? Our suggestion is that it contributes something very like a presupposition about the addressee, just as second-person pronouns do. For an addressee term, the presupposition has additional content as well, namely, all the properties which it mentions. For instance, here is an extract from Example 10 with heightened passion:

(15) Bob: I love you, Sue.

The DRS for this example is given in Figure 3. There are two addressee pre-

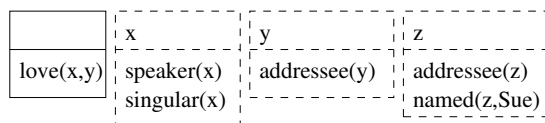


Fig. 3. DRS for *I love you, Sue*

suppositions here, one contributed by *you*, and one by *Sue*. Note that the latter presupposition does not bind to anything in the assertion DRS, but that because there is only ever one addressee entity in the context, the two addressee presuppositions are constrained to corefer in this case.

There are nonetheless some important differences between the semantics of an addressee term and that of a second-person pronoun. Firstly, a plural addressee term squarely presupposes the addressee entity, not simply an entity including the addressee. Secondly, as emphasised in Section 3.2, addressee terms can function to *change* the default addressee, not simply to refer. This process needs to be described in some detail.

Assume we are dealing with a new incoming utterance from the user. Before processing this utterance, the system the system deletes all the existing **speaker** and **addressee** predicates from the context, sets the new **speaker** to be the user, and sets the **addressee** predicate according to the defaults given in Section 3. For instance, consider Example 16:

(16) User: I like movies, Bob.
 Bob (played by the system): That's interesting.
 User: Do you like movies, Sue?

After Bob’s utterance, the system will set **speaker** to User, and **addressee** to Bob, by Principle 4. The user’s second utterance presupposes that Sue is the addressee. The only way of allowing this is by accommodation of an **addressee** predicate which holds of the object named Sue. Since there can only ever be one **addressee** predicate, this also means deleting the existing **addressee** predicate holding of Bob. Accommodation of presuppositions due to addressee terms is thus non-monotonic. One problem this raises is in how to handle the presupposition due to the pronoun *you* in the utterance. If it was processed before that of the addressee term, it ends up being bound to an object which is no longer the addressee! These complications make it important to handle presuppositions due to addressee terms before any other presuppositions. We therefore augment the DRS language we use to represent utterances, to include a special sub-DRS for addressee terms.

6 An example dialogue for practicing pronouns in conversational Māori

The main application we have in mind for our dialogue system is a computer-aided language learning system for Māori which uses dialogue as its paradigm. To teach the Māori pronoun system, it is useful to support more than two speakers. We have implemented all of the syntax and semantics of pronouns and addressee terms, and many of the dialogue management principles, but we still have some work to do in combining these into a unified framework. In the meantime, we give an example below of the kind of multi-speaker dialogue we have in mind. The student is Jason; Hone and Piri are characters whose utterances are generated by the dialogue system.

1	Hone/Piri	Kia ora, Jason!	Hello, Jason!
2	Jason	Tēna kōrua.	Hello (you two).
3	Hone	Kei te hiakai ahau.	I’m hungry.
4	Piri	Kāore ahau i te hiakai.	I’m not hungry.
5	Piri	Kei te hiakai koe, Jason?	Are you hungry, Jason?
6	Jason	Kāore.	No.
7	Hone	Jason, kei te hiakai kōrua ko Piri?	Jason, are you (dual) and Piri hungry?
8	Jason	Kāore māua i te hiakai.	No, we (dual, exclusive) aren’t hungry.

In Utterance 1, Hone and Piri are the speakers. They address Jason explicitly to start with. When Jason responds, Hone and Piri are by default the addressees, and Jason’s greeting uses the appropriate dual pronoun. In 3 and 4, Hone and Piri provide some information about themselves. In 5, Piri asks a question of Jason. Since Hone is the default addressee of this utterance, Piri has to identify Jason explicitly with an addressee term. In 6, Jason answers Piri. In 7, Hone jumps in; by default, his utterance therefore has Jason and Piri as its addressees. However, Hone’s utterance has an explicit addressee term (*Jason*) which overrides this. Finally, in 8, Jason replies, using an appropriate dual and exclusive first-person pronoun to refer to himself and Piri. The utterance is assumed to be addressed to Hone by default.

Implementing a dialogue system which supports this kind of interaction will allow a student great flexibility in practicing Māori questions, answers, assertions, all in a context where distinctions between the alternative personal pronouns are clearly motivated.

References

1. Padilha, E., Carletta, J.: A simulation of small group discussion. In: Proceedings of the 6th workshop on the semantics and pragmatics of dialogue (EDILOG 2002), Edinburgh (2002) 117–124
2. Wang, H.C., Huang, C.Y., Yang, C.H., Wang, J.F.: Multi-speaker dialogue for mobile information retrieval. In: Proceedings of the international symposium on Chinese spoken language processing (ISCSLP), National Cheng-Kung University, Tainan, Taiwan (2002)
3. de Jager, S., Knott, A., Bayard, I.: A DRT-based framework for presuppositions in dialogue management. In: Proceedings of the 6th workshop on the semantics and pragmatics of dialogue (EDILOG 2002), Edinburgh (2002)
4. Bayard, I., Knott, A., de Jager, S.: A unification-based grammar for a fragment of Māori and English. In: Proceedings of the 2nd Australasian Natural Language Processing Workshop (ANLP 2002). (2002)
5. Copestake, A.: The (new) LKB system. CSLI, Stanford University (2000)
6. Knott, A., Vlugter, P.: Syntactic disambiguation using presupposition resolution. In: Proceedings of the 4th Australasian Language Technology Workshop (ALTW2003), Melbourne (2003)
7. Copestake, A., Flickinger, D., Sag, I., Pollard, C.: Minimal Recursion Semantics: An introduction. Manuscript, CSLI, Stanford University (1999)
8. Kamp, H., Reyle, U.: From discourse to logic. Kluwer Academic Publishers, Dordrecht (1993)