

# Generating Feedback and Sequencing Moves in a Dialogue System

**Staffan Larsson**

Department of Linguistics  
Göteborg University  
Box 200  
SE40530 Göteborg  
Sweden  
sl@ling.gu.se

## Abstract

In current dialogue systems, user utterances are often verified by the system to make sure the system heard the user right. While this mechanism is useful and important in dialogue systems, it covers only a small part of the spectrum of interactive communication management (ICM) devices occurring in human-human dialogue. A view of the full range of ICM dialogue moves can be useful for suggesting useful additions to the range of moves handled by dialogue systems, both by making systems more flexible in terms of dialogue management and by making dialogues more natural. Starting from a fairly broad categorization of various types of feedback and sequencing moves in human-human dialogue, we select and implement a subset of these moves in a dialogue system.

## Introduction

In this paper<sup>1</sup>, we will give an account of the generation of feedback and sequencing in a dialogue system. Allwood (1995) uses the concept of “Interactive Communication Management” (ICM) to designate all communication dealing with the management of dialogue interaction. This includes feedback but also *sequencing* and *turn management*. Sequencing “concerns the mechanisms, whereby a dialogue is structured into sequences, subactivities, topics etc. ...”.

Here, we will use the term ICM as a general term for coordination of the common ground, which in an information state update approach comes to mean explicit signals enabling coordination of updates to the common ground. While feedback is associated with specific utterances, ICM in general does not need to concern any specific utterance. We will also be making use of various other parts of Allwood’s “activity-based pragmatics” (Allwood 1995), including Allwood’s action level terminology, the concept of Own Communication

Management (OCM), and various distinctions concerning ICM.

We start by briefly discussing ICM in current dialogue systems. We then proceed to give an overview of ICM, and especially feedback, in human-human communication. Finally, we select a limited range of ICM behaviour which we consider useful for dialogue systems, and show a brief example of how these utterances are selected and generated in GoDiS (Larsson 2002), a dialogue system based in issue-based dialogue management and implemented using the TrindiKit (Traum & Larsson forth).

## Verification

In the literature concerning practical dialogue systems, e.g. San-Segundo *et al.* (2001), grounding is often reduced to verification of the system’s recognition of user utterances. Two common ways of handling verification are described as “explicit” and “implicit” verification, exemplified below (example from San-Segundo *et al.* (2001)).

- I understood you want to depart from Madrid. Is that correct? [explicit]
- You leave from Madrid. Where are you arriving at? [implicit]

Actually, both “explicit” and “implicit” feedback contain a verbatim repetition or a reformulation of the original utterance, and in this sense they are both explicit. The actual base for the distinction is that the first utterance (1) signals understanding but also a lack of confidence in the interpretation and (2) is trying to elicit a response regarding the correctness of the interpretation from the hearer, whereas the second utterance (1) signals confident understanding and (2) does not try to elicit a response.

In human-human dialogue, explicit confirmations occur in noisy environments and in situations where understanding is critical (e.g. when arranging a meeting in a busy airport). Given that verification is presumably a rather marginal phe-

<sup>1</sup>This paper is based on chapter 4 of Larsson (2002).

nomina in human-human dialogue, it is perhaps surprising that it is often the only aspect of feedback covered in dialogue systems literature. Firstly, because it is usually not necessary for humans to verify what they (think they) have heard; that is, it is a rather uncommon grounding procedure in human-human dialogue. Second, because it only involves part of the full spectrum of feedback behaviour, excluding e.g. acceptance-related feedback behaviour.

Of course, verification of user utterances are of central importance in dialogue systems, given the quality of current speaker-independent speech recognition. This explains to some extent why verification is often the only aspect of feedback handled by current systems - it is simply necessary. However, this is no reason not to explore further the possible uses of a wider range of feedback behaviour in dialogue systems.

### Feedback and related behaviour in human-human dialogue

By feedback we mean behaviour whose primary function is to deal with grounding of utterances in dialogue<sup>2</sup>. This distinguishes feedback from behaviour whose primary function is related to the domain-level task at hand, e.g. getting price information. Non-feedback behaviour in this sense includes asking and answering task-level questions, giving instructions, etc. (cf. the “Core Speech Acts” of Poesio & Traum (1998)). Answering a domain-level question (e.g. saying “Paris” in response to “What city do you want to go to?”) certainly involves aspects of grounding and acceptance, since it shows that the question was understood and accepted. However, the primary function of a domain-level answer is to resolve the question, not to show that it was understood and accepted.

To get an overview of the range of explicit feedback behaviour that exists in human-human dialogue, we will classify feedback according to four criteria. We will assume that DP *S* has just uttered or is uttering *u* to DP *H*, when the feedback utterance *f* (uttered by *H* to *S*) occurs.

- level of action / basic communicative function (contact, perception, understanding, reaction / acceptance)
- polarity (positive / negative / neutral): whether *f* indicates contact / perception / understanding / acceptance or lack thereof
- eliciting / non-eliciting: whether *f* is intended to evoke a response (e.g. a reformulation or a reason to accept some content)
- form of *f*: single word, repetition etc.
- content of *f*: object-level or meta-level

<sup>2</sup>Since this paper is not concerned with multimodal dialogue, we will only discuss verbal feedback.

The action level criterion has been explained above; the others will be explained presently. The criteria of basic communicative function, polarity, eliciting/non-eliciting, and surface form are all derived from Allwood, Nivre, & Ahlsen (1992) and Allwood (1995).

### Levels of action in dialogue

Both Allwood (1995) and Clark (1996) distinguish four levels of action involved in communication (*S* is the speaker of utterance *u*, *H* is the hearer/addressee). They use slightly different terminologies; here we use Allwood’s terminology and add Clark’s (and, for the reaction level, also Ginzburg’s) corresponding terms in parenthesis. The definitions are mainly derived from Allwood.

- Reaction (acceptance, consideration): whether *H* has integrated (the content of) *u*
- Understanding (recognition): whether *H* understands *u*
- Perception (identification): whether *H* perceives *u*
- Contact (attention): whether *H* and *S* have contact, i.e. if they have established a channel of communication

Given that grounding is concerned with all levels, it follows that grounding-related feedback may concern any (and possibly several) of these levels.

We can make further distinctions between different levels of understanding, corresponding to three levels of meaning. These sublevels give a finer grading to the level of understanding. (A similar distinction is also used by Ginzburg (forth)).

- domain-dependent and discourse-dependent meaning (roughly, “content” in the terminology of Barwise & Perry (1983) and Kaplan (1979))
  - referential meaning, e.g. referents of pronouns, temporal expressions
  - pragmatic: the relevance of *u* in the current context
- discourse-independent (but possibly domain-dependent) meaning (roughly corresponding to “meaning” in the terminology of Barwise & Perry (1983) and Kaplan (1979)), e.g. static word meanings

By “discourse-independent” we mean “independent of the dynamic dialogue context” (modeled in GoDiS by the information state proper). However, discourse-independent meaning may still be dependent on static aspects of the activity/domain. It is obvious that these levels of meaning are intertwined and do not have perfectly clear boundaries. Nevertheless, we believe they are useful as analytical approximations.

## Positive, negative, and neutral feedback

Positive feedback indicates one or several of contact, perception, understanding, and integration, while negative feedback indicates lack thereof.

While there are clear cases of positive (“uhuh”, “ok”) and negative (“pardon?”, “I don’t understand”) feedback, there are also some cases which are not so clear. For example, are check-questions (e.g. “To Paris?” in response to “I want to go to Paris”) positive or negative? If positive feedback shows understanding, and negative feedback lack of understanding, then check-questions are somewhere in between; they indicate understanding but also that the lack of confidence in that understanding.

Here we will assume a third category of *neutral* feedback for check-questions and similar feedback types. If negative feedback indicates a lack of understanding, neutral feedback indicates lack of confidence in one’s understanding.

Negative feedback can be caused by failure to integrate *U* on any of the levels of action in dialogue:

- lack of contact - H did not notice that S said something
- lack of perception - H did not hear what S said
- lack of understanding on a semantic/pragmatic level - H recognized all the words, but could not extract a content
  - context-independent meaning, e.g. word meanings
  - context-dependent meaning, e.g. referents
  - pragmatic meaning, i.e. the relevance of S’s utterance in relation to the context
- rejection of content

## Eliciting and non-eliciting feedback

We will use the term “eliciting feedback”, borrowed from (Allwood, Nivre, & Ahlsen 1992), to refer to feedback utterances intended to elicit a response, or more specifically utterances *u'* such that *u'* is intended to make *S* respond to *u'* because *H* is not sure about how to interpret *S*’s utterance *u*. Check-questions (both *y/n*- and alternative-questions) are seen as eliciting feedback in this sense. Eliciting feedback can also occur after *S*’s utterance *u* is finished.

## Form of feedback

As with all utterances, feedback utterances can have various syntactic forms:

- declarative (“I heard you say ‘go to Paris.’”, “You want to go to Paris.”)
- imperative (“Please repeat.”)

- interrogative
  - *y/n*-question (“Did you say ‘Paris’?”, “Do you want to go to Paris?”)
  - *wh*-question (“What did you say?”, “What do you mean?”, “Where do you want to go?”)
  - alternative-question (“Did you say ‘Paris’ or ‘Ferris’?”, “Do you want to go to Paris, France or Paris, Texas?”)
- ellipsis (“Paris?”, “to Paris.”)
- conventional (“Pardon?”)

Apart from showing the speaker that one has understood, feedback in the form of an explicit declarative report, repetition or reformulation has the additional function of making sure that the understanding is actually correct, by providing a chance for correction. A *y/n*-question has a similar function, but it indicates less confidence in the interpretation (i.e. is more neutral) and has a stronger eliciting element than an assertion; a question requires an answer, while an assertion can often be assumed to be accepted in the absence of protest.

A related dimension of classification is how the form of the feedback utterance relates to the previous utterance. One way of giving positive feedback is to simply repeat verbatim the previous utterance (e.g. “To Paris.” in response to “To Paris.”). A similar strategy is to provide a reformulation (e.g. “Your destination city is Paris, the capital of France.”). The latter is perhaps a stronger signal of understanding than the former, since a verbatim repetition does not in principle require that the utterance was understood.

## Meta-level and object-level feedback

A final distinction can be made depending on whether the feedback explicitly talks about what the speaker said or meant, in which case the feedback can be said to be *meta-level* feedback, or if instead it talks about the subject matter of the dialogue, in which case we talk about *object-level* feedback.

- Meta-level
  - perception (“Did you say ‘Paris’?”)
  - understanding (“Did you mean that you want to go to Paris?”)
- Object-level (“Do you want to go to Paris?”)

This distinction does not necessarily apply to all kinds of feedback. For example, for conventional phrases like “Pardon?” and elliptical phrases like “Paris?” it is not clear if they refer to what the speaker said or meant, or about the subject matter of the dialogue, or neither.

## Feedback in GoDiS

In this section, we first show how feedback dialogue moves in GoDiS are represented. We then review the full range of feedback moves, starting with system-generated feedback and then moving on to user feedback.

The general notation for ICM dialogue moves used in GoDiS is the following:

- $icm:L * P\{ :Args \}$

where  $L$  is an action level,  $P$  is a polarity, and  $Args$  are arguments.

- $L$ : action level
  - con: contact (“Are you there?”)
  - per: perception (“I didn’t hear anything from you”, “I heard you say ‘to Paris’”)
  - sem: semantic understanding (“I don’t understand”, “To Paris.”)
  - und: pragmatic understanding (“I don’t quite understand”, “You want to know about price.”)
  - acc: acceptance/reaction (“Sorry, I can’t answer questions about connecting flights”, “Okay.”)
- $P$ : polarity
  - neg: negative
  - int: interrogative
  - pos: positive
- $Args$ : arguments

Note that the “neutral” polarity has been replaced by the label “int”; we have made a simplifying assumption that neutral feedback is always eliciting and interrogative.<sup>3</sup>

The arguments are different aspects of the utterance or move which is being grounded, depending action level:

- for per-level: *String*, the recognized string
- for sem-level: *Move*, a move interpreted from the utterance
- for und-level:  $DP * C$ , where
  - $DP$  : Participant is the DP who performed the utterance
  - $C$  : Proposition is the propositional content of the utterance
- for acc-level:  $C$  : Proposition, the content of the utterance

<sup>3</sup>Note, however, that if we had included feedback forms like “What did you say?”, this would still be regarded as negative feedback. The “int” label only refers to check-questions, which are usually *y/n*-questions. This is arguably not an optimal labelling convention.

For example, the ICM move  $icm:und*pos:usr*dest-city(Paris)$  provides positive feedback regarding a user utterance that has been understood as meaning that the user wants to go to Paris.

## System feedback to user utterances in GoDiS

The fact that speech recognition is much harder for machines than for humans may motivate different grounding strategies for handling system utterances than for handling user utterances. In GoDiS, the focus is on the system’s capabilities for generating feedback in reaction to user utterances. However, the system is also able to react appropriately to a limited range of user feedback, as will be seen in the next section.

For user utterances, GoDiS will be able to produce e.g. the following kinds of feedback utterances (for the examples, assume that the user just said “I want to go to Paris”):

- contact
  - negative;  $icm:con*neg$  (“I didn’t hear anything from you”)
- perception
  - negative;  $icm:per*neg$  realized as fb-phrase (“Pardon?”, “I didn’t hear what you said.”)
  - positive;  $icm:per*pos:String$  realized as metalevel verbatim repetition (“I heard ‘to paris’ ”)
- understanding (semantic)
  - negative;  $icm:sem*neg$  realized as fb-phrase (“I don’t understand.”)
  - positive;  $icm:sem*pos:Content$  realized as repetition/reformulation of content (object-level) (“Paris.”)
- understanding (pragmatic)
  - negative;  $icm:und*neg$  realized as fb-phrase (“I don’t quite understand.”)
  - positive;  $icm:und*pos:DP*Content$  realized as repetition/reformulation of content (object-level) (“To Paris.”)
  - interrogative;  $icm:und*int:DP*Content$  realized as ask about interpretation (“To Paris, is that correct?”)
- integration
  - negative
    - \* proposition-rejection;  $icm:acc*neg:Content$  realized as explanation (“Sorry, Paris is not a valid destination city”)
  - positive;  $icm:acc*pos$  realized as fb-word (“Okay”)

In addition, GoDiS will be able to reject user questions using the  $icm:acc*neg:issue(Q)$  move, where  $Q$  : Question as illustrated below:

U> What about connecting flights?

S> Sorry, I cannot answer questions about connecting flights.

We are not claiming that humans always make these distinctions between action explicitly or even consciously, nor that the link between surface form and feedback type is a simple one-to-one correspondence; for example, “mm” may be used as positive feedback on the perception, understanding, and acceptance levels. Feedback is, simply, often ambiguous. However, since GoDiS is making all these distinctions internally we might as well try to produce feedback which is not so ambiguous. Of course, there is also a tradeoff in relation to brevity; extremely explicit feedback (e.g. “I understood that you referred to Paris, but I don’t see how that is relevant right now.”) could be irritating and might decrease the efficiency of the dialogue. We feel that the current choices of surface forms are fairly reasonable, but testing and evaluation on real users would be needed to find the best ways to formulate feedback on different levels. This is an area for future research.

A general strategy used by GoDiS in ICM selection is that if negative or interrogative feedback on some level is provided, the system should also provide positive feedback on the level below. For example, if the system produces negative feedback on the pragmatic understanding level, it should also produce positive feedback on the semantic understanding level.

### User feedback to system utterances in GoDiS

For system utterances, GoDiS will react appropriately to the following types of user feedback:

- perception level
  - negative; fb-phrase (“Pardon?”, “Excuse me?”, “Sorry, I didn’t hear you”) interpreted as `icm:per*neg`
- reaction/acceptance level
  - positive; fb-phrase (“Okay.”) interpreted as `icm:acc*pos`
  - negative; issue rejection fb-phrase (“I don’t know”, “Never mind”, “It doesn’t matter”) interpreted as `icm:acc*neg:issue`

In addition, irrelevant followups to system ask-moves are regarded as implicit issue-rejections. The coverage of user feedback behaviour is thus more limited than the coverage for system behaviour. The main motivation for this is that system utterances are less likely to be problematic for the user to interpret than vice versa.

## Sequencing ICM in GoDiS

We believe it is good practice to try to keep the user informed about what’s going on inside the system, at least to a degree that facilitates a natural dialogue where system utterances “feel natural”. One way of doing this is to produce ICM phrases indicating significant updates to the information state which are not directly related to specific user utterances.

For GoDiS, we have implemented three types of sequencing ICM. First, GoDiS will indicate when it is loading a dialogue plan<sup>4</sup>. Second, GoDiS will produce ICM to indicate when an issue is being reraised (in contrast to being raised for the first time). Third, GoDiS generates ICM moves for indicating accommodation of questions (similar to plan recognition). This is meant to indicate that some inference has been performed, and that it has been successful.

The choice of actual wordings for these moves are currently only based on intuitions; both corpus studies and usability tests can serve to give a firmer ground to the exact formulation of sequencing moves.

- `icm:reraise`: indicate reraising implicitly (“So, ...”)
- `icm:reraise:Q`: reraising an issue *Q* explicitly (“Returning to the issue of Price.”)
- `icm:loadplan` (“Let’s see.”)
- `icm:accommodate` (“Alright.”)
- `icm:accommodate:Q` (“Alright, you want to know about ...”)

### A sample dialogue with GoDiS

The dialogue below gives an impression of what feedback and sequencing capabilities our system handles. The basic principle for move selection in GoDiS is that utterances may not raise more than one question. Surface form generation in GoDiS is template-based.

The dialogue has been annotated with ICM dialogue moves. Although it lies outside the scope of this paper to present the formalism used in GoDiS for representing semantic content, it has been included in the ICM move annotation for completeness. We hope the formalism is sufficiently straightforward to be understood by the reader; if not, the content-parts of the annotation can be ignored.

S> Welcome to the travel agency!

U> price information please

S> Okay. [`icm:acc*pos`] You asked about price.

<sup>4</sup>Due to space limitations we will not here go into the internal workings which trigger sequencing ICM. See (Larsson 2002) for detailed accounts of these.

[icm:und\*pos:usr\*issue(?x.price(x))]. Let's see.  
 [icm:loadplan] How do you want to travel?  
 U> pardon [icm:per\*neg]  
 S> How do you want to travel?  
 U> by plane, london  
 S> By plane, is that correct?  
 [icm:und\*int:transport(plane)]  
 U> yes  
 U> London. [icm:sem\*pos:london] I don't quite understand. [icm:und\*neg] Do you mean to London or from London?  
 U> to london  
 S> Okay [icm:acc\*pos], to London [icm:und\*pos:dest-city(london)]. What city do you want to go from?  
 U> lyons  
 S> Sorry, Lyons is not in the database [icm:acc\*neg:dept-city(lyons)]. So, [icm:reraise] What city do you want to go from?  
 U> paris  
 S> Okay [icm:acc\*pos], to Paris [icm:und\*pos:dept-city(paris)]. When do you want to travel?  
 U> Do I need a Visa?  
 S> Okay [icm:acc\*pos], you want to know about visa [icm:und\*pos:issue(?need-visa)]. What country are you from?  
 U> Sweden  
 S> Okay, from Sweden. No, you don't need a visa. Returning to the issue of price [icm:reraise:?x.price(x)]. So, [icm:reraise] when do you want to travel?  
 ...  
 U> What about connecting flights?  
 S> Sorry, I cannot answer questions about connecting flights. [icm:acc\*neg:issue(?x.connecting-flights(x))]

## Conclusions and future research

Starting from a fairly broad categorization of various types of feedback and sequencing moves in human-human dialogue, we have implemented a subset of these moves in a dialogue system. We believe that a view of the full range of ICM behaviour can be useful for suggesting useful additions to the range of moves handled by dialogue systems, both by making systems more flexible in terms of dialogue management and by making dialogues more natural.

Our account of grounding and ICM is so far only partial in coverage; phenomena that remain to be accounted for and/or implemented include clarification ellipsis, semantic ambiguity resolution, collaborative completions and repair, and turntaking ICM. While we have included some rudimentary sequencing ICM, further investigations of the appropriateness and usefulness of these utterances are needed; here, research on discourse markers (e.g. Schiffrin (1987)) and cue phrases (e.g. Grosz & Sidner (1986), Polanyi & Scha (1983), and Reichman-Adar (1984)) can be of great use. We also want to explore turntaking in asynchronous dialogue management, and how this relates to turntaking ICM.

We are currently implementing a VCR control application for GoDiS to enable evaluation and further development of the system. This application will be reachable by phone and allow users to program and control a (computerized) VCR using spoken dialogue.

## References

- Allwood, J.; Nivre, J.; and Ahlsen, E. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics* 9:1–26.
- Allwood, J. 1995. An activity based approach to pragmatics. Technical Report (GPTL) 75, Gothenburg Papers in Theoretical Linguistics, University of Göteborg.
- Barwise, J., and Perry, J. 1983. *Situations and Attitudes*. The MIT Press.
- Clark, H. H. 1996. *Using Language*. Cambridge: Cambridge University Press.
- Ginzburg, J. forth. Questions and the semantics of dialogue. Forthcoming book, partly available from <http://www.dcs.kcl.ac.uk/staff/ginzburg/papers.html>.
- Grosz, B. J., and Sidner, C. L. 1986. Attention, intention, and the structure of discourse. *Computational Linguistics* 12(3):175–204.
- Kaplan, D. 1979. Dthat. In Cole, P., ed., *Syntax and Semantics v. 9, Pragmatics*. New York: Academic Press. 221–243.
- Larsson, S. 2002. *Issue-based Dialogue Management*. Ph.D. Dissertation, Göteborg University.
- Poesio, M., and Traum, D. R. 1998. Towards an axiomatization of dialogue acts. In *Proceedings of Twendial'98, 13th Twente Workshop on Language Technology: Formal Semantics and Pragmatics of Dialogue*, 207–222.
- Polanyi, L., and Scha, R. 1983. On the recursive structure of discourse. In Ehlich, K., and van Riemsdijk, H., eds., *Connectedness in Sentence, Discourse and Text*. Tilburg University. 141–178.
- Reichman-Adar, R. 1984. Extended man-machine interface. *Artificial Intelligence* 22(2):157–218.
- San-Segundo, R.; Montero, J. M.; Guitierrez, J. M.; Gallardo, A.; Romeral, J. D.; and Pardo, J. M. 2001. A telephone-based railway information system for spanish: Development of a methodology for spoken dialogue design. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, 140–148.
- Schiffrin, D. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.
- Traum, D., and Larsson, S. forth. The information state approach to dialogue management. In Smith, R., and Kuppevelt, J., eds., *Current and New Directions in Discourse & Dialogue*. Kluwer Academic Publishers.