Non-problems and social obligations in human-computer conversation

Harry Bunt
Tilburg University
PO Box 90153, 5000 LE Tilburg, NL
bunt@kub.nl

1 Introduction

To be truly conversation-like, a human-computer dialogue has to have much of the freedom and flexibility of human-human conversation. This means that a conversational dialogue system should be able to handle a wide range of communicative acts that people commonly employ in dialogue. It should be able both to understand these acts when they are used by the human partner, and to use them acts in situations where the human partner would expect this. In this paper we report on a comparative study of human-human and WOZ-simulated human-computer dialogues concerned with the same information-seeking task, investigating the differences in the use of dialogue act types, and we discuss the implementation of the dialogue act types that seem to be needed in conversational dialogue systems.

2 Theoretical background

We use a theoretical framework called Dynamic Interpretation Theory (DIT) (see Bunt, 1989; 1994; 1999; 2000), in which communicative acts, or *'dialogue acts'* are defined (a) semantically in terms of context-change potential, but also (b) linguistically in terms of utterance features, in that a certain context changing action is viewed as a communicative act only if there are linguistic or or paralinguistic aspects of an agent's communicative behaviour that signal the intention to change the context in this particular way.[1] With this definition, we have developed a corpus-based taxonomy of dialogue act types, whose top level makes a distinction between those types of acts that serve to accomplish something in the underlying task, motivating the dialogue, and those types of acts that serve to create and maintain the conditions for smooth and successful interaction. Dialogue acts of the first kind we call *task-oriented (TO) acts*; those of the latter kind we call *dialogue control (DC) acts* (see Figure 1).

TO-acts have subdivisions that are of no concern here (see e.g. Bunt, 1999); DC acts are subdivided into *feedback* (FB) acts, *interaction management* (IM) acts, and *social obligations management* (SOM) acts.

Feedback acts are subdivided into negative FB acts, that deal with some communicative or information processing problem (like incorrect understanding), and positive ones, that indicate the absence of such problems (acknowledgements, for instance). IM acts

[1] What we mean by 'paralinguistic' here depends on the communicative modalities that are used, e.g. intonation and speech tempo, or use of capitals and underlining, or use of highlighting and beeps, or use of accompanying gestures and mimics.

are similarly subdivided into those that deal with some problem or uncertainty in the interactive situation (like whose turn it is to say something, or whether there is perceptual contact), and anticipatory ones that serve to prevent problems of this sort, like *"Just a moment please"*. Both kinds of FB acts and IM acts have further subdivisions (see Bunt, 1999; 1996). SOM acts are subdivided in those concerned with greeting, introducing oneself, thanking, apologizing, and valediction.

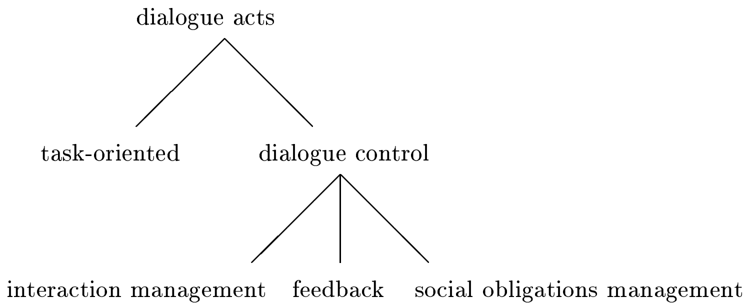

Figure 1: *Types of dialogue acts.*

According to Dynamic Interpretation Theory, a dialogue act consists semantically of two components, called *communicative function* and *semantic content* (for short also 'function' and 'content'). The semantic content of a dialogue act is the new information that the act introduces in the current context; the communicative function specifies in what way the speaker intends to update the context with the content. Formally, given a context Γ and a dialogue act A with function F and content p, we have $(F(p))(\Gamma) = \Gamma'$, where Γ' is the context after the act A has been performed successfully; the *meaning* of the dialogue act is the set of transitions $\{< \Gamma, \Gamma' > |(F(p))(\Gamma) = \Gamma'\}$. Besides a formal semantic structure, a dialogue act also has an *utterance form* with features that allow a speaker to recognize the act, in particular its communicative functions. (Though of course the features of an utterance are most of the time insufficient to determine its functions unambiguously.) Figure 2 shows part of the top of the hierarchical system of communicative functions that we have distinguished.

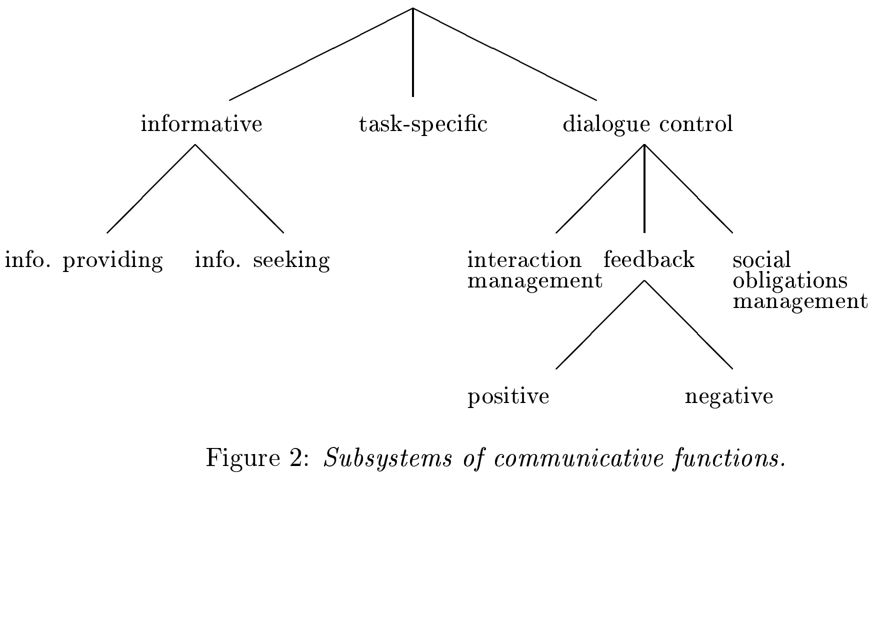

Figure 2: *Subsystems of communicative functions.*

The hierarchy of communicative functions has a 3-way rather than a 2-way split at the top, since general information providing and -seeking functions (the various kinds of questions, statements and answers) can be used both to build task-oriented and to construct dialogue control acts, depending on the semantic content with which they are combined. We will return to this below.

An important point about the relation between utterances and dialogue acts is that utterances are often *multifunctional*, having multiple communicative functions. For instance, an utterance may answer a question and at the same time provide positive feedback and give the turn to the dialogue partner.

3 Dialogue act types in human-computer dialogue

The kind of task-oriented dialogue acts that one finds in a human-human dialogue are determined by the kind of underlying task; in information-seeking dialogues one finds e.g. questions, answers, and requests; in a negotiation dialogues one finds offers, accepts and refusals; in a tutoring dialogue one finds explanations, corrections and verifications (see e.g. Hulstijn, 2000). It seems rather obvious that human-computer dialogues will also contain the TO acts that we find in human-human dialogues concerned with the same task, because these acts are by definition appropriate for that task.

Of the various categories of dialogue control acts, in addition, we clearly need negative FB acts and problem-oriented IM-acts for dealing with misunderstandings and other communicative problems, which arise more easily with a machine then in human-human communication. Indeed, dialogue system designers often focus on these two categories of dialogue control, often referred to as 'error handling'. In human-human dialogues we find, in addition, a great deal of positive feedback acts and anticipatory IM acts, as well as acts that seem to be mainly motivated by the fact that communication is a social activity: SOM acts such as greeting, thanking, and apologizing, which all come in initiative-response pairs, like *"I'm sorry"/"That's all right"* and *"Merci"/"Pas de quoi"* (cf. Bilange, 1991; Moeschler, 1989).

The study reported here is specifically concerned with these 'non-problem' FB- and IM acts, and with socially motivated acts. For acts in these categories it is not so obvious to what extent we need or we want them in human-computer dialogues, and if we do, what mechanisms underlying the use of such acts can be implemented in a dialogue management system.

Reeves and Nass (1996) report a number of studies suggesting that people in their interactive behaviour with computers (unconsciously) treat these in many respects as social actors. They argue in particular that politeness is important in human-computer interaction, playing a greater role than one would expect, and that people, both novice and expert users, have a natural tendency to interact with computers in a polite way, not very different from when they interact with a human conversation partner. The results they report are insufficiently detailed to say much about the use of specific types of SOM acts like those mentioned above: greeting, thanking, introducing oneself, apologizing and valediction. Still less clear is whether the tendency to behave towards computers as social actors means that 'non-problem FB- and IM acts' are also used in much the same way as when interacting with a human dialogue partner. The experiment described in the next section sheds some more light on these issues.

4 Experimental investigation

We analysed a corpus of 192 information-seeking dialogues collected in an experiment that was originally intended to explore the influence of various details of keyboard functionality on the communicative behaviour of dialogue participants. The dialogues were conducted using computer keyboards and screens in four different conditions. In three conditions subjects were told they were communicating with a computer (with slightly different technical properties), which was in fact simulated by a wizard; the fourth condition corresponded functionally to the telephone information service at Schiphol, Amsterdam Airport. The information-seeking scenarios used in the experiment were taken from recorded telephone dialogues with the airport information service. All dialogues were transcribed and annotated with the communicative functions defined in Bunt (1999) and Bunt (2000), and analysed for the occurrences of 'non-problem FB- and IM acts' and SOM acts.

For the occurrence of SOM acts the results are summarized in Table 1. The first two columns contain the numbers of occurrences in the 144 human-computer dialogues, distinguishing between use by the human partner and the computer; the next two columns concern the use of SOM acts by the client and the information service in the 48 human-human dialogues.

	HC (144)		HH (48)	
SOM act type	human	computer	client	info. service
welcome greeting, initiative	14	1	20	0
welcome greeting, reaction	1	3	0	10
self-introduction	9	1	7	4
thanks (initiative)	50	0	30	0
thanks reaction	0	9	0	29
apology (initiative)	0	2	1	1
apology reaction	0	0	0	0
valediction, initiative	12	2	6	0
valediction, reaction	0	2	0	5

Table 1. Occurrence of social obligation management acts.

It should be noted that the wizard in the HC conditions was instructed to behave in a very sober fashion, avoiding saying anything that would not seem strictly necessary to say in order to enhance the feeling of interacting with a machine. This is reflected in the low numbers of SOM acts by the wizard in Table 1, and also results in low numbers of positive FB-acts and anticipatory IM-acts. Anticipatory IM-acts consist of four main types: those concerned with time management (*"Just a moment please"*), those concerned with discourse structuring (*"One other question"*), those concerned with contact management (*"Hello"*), and turn management acts where one participant interrupts the other. Table 2 summarizes the results, using '-' for 'does not occur'; '+' for 'occurs occasionally'; '++' for 'occurs fairly commonly', and '+++' for 'occurs frequently'.

	HC		HH	
dialogue act type	human	computer	client	info. service
positive feedback	++	+	+++	+++
time management	-	++	+	++
discourse structuring	++	-	++	-
contact management	++	+	++	++
interrupting	-	+	+	+

Table 2. Occurrence of positive feedback and anticipatory IM acts

5 Discussion and conclusions

The tables 1 and 2 show that, with the exception of reactions to apologies, every type of SOM act, positive FB act, and anticipatory IM act is used in the human-computer dialogues either by the user or by the wizard (or by both). Hence the general conclusion seems warranted that a conversational dialogue system should be able to use and understand dialogue acts of all the categories distinguished here. There is a clear tendency that positive feedback and social obligations management occur more frequently in human-human than in human-computer interaction, which casts some doubts on Reeves and Nass' claim that people treat computer dialogue partners in essentially the same way as human dialogue partners, but perhaps more important is the conclusion that all dialogue acts that might be thought to be characteristic for human dialogue, like greeting, thanking and saying goodbye, do occur also in human-computer dialogues, and are therefore of potential relevance for the design of computer dialogue systems. It should also be noted that these results were obtained for typed dialogues, using keyboards, and that the use of speech would almost certainly lead to more frequent use of these dialogue acts.

The experiment also tells us something about the internal structure of the SOM acts and the non-problem FB- and IM acts in human-computer dialogues. All dialogue control acts can be constructed in two ways: (1) by using a specific communicative FB-, IM-, or SOM function combined with marginal semantic content, or (2) by using a general information-providing or -seeking function (i.e., various types of assertions and questions; see Figure 2) in combination with a full-blown propositional content. This difference is reflected in the surface form of the utterances used in each case, e.g. *"Thanks"* is a SOM act with construction (1), and *"I'm very grateful for your help"* is one with construction (2).

In the experiment we find that only positive FB acts, anticipatory IM acts, and SOM acts with construction (1) occur in human-computer dialogues, whereas the human-human dialogues also contain substantial numbers of acts with construction (2). For instance, 6 of the 30 expressions of thanks and 5 of the 6 valedictions in the human-human dialogues have construction (2). This phenomenon can be explained by considering the context-changing effects associated with communicative functions. When a dialogue control act is constructed using a general informative function and a full-blown semantic content, then the utterance that is used to perform the act has the form of a sentence that is in general used to express a statement, question, or answer, like the example *"I'm very grateful for your help"*, which has the form of a statement. Such a sentence indeed expresses in the first place that the sender informs the receiver that he, the sender, is grateful for the receiver's help; the sender wishes to change the context to the effect that the hearer now knows this to be the case. This may be contrasted with the use of the standard formula *"Thanks"*, where the sender is not so much intending the make the receiver know that he is grateful, but is rather dealing with a social obligation that arises when some point is reached in the interaction. (And it changes the context to the extent that it draws the receiver's attention to the fact that this is the case.) In human-computer interaction, it seems pointless to try to convince a computer that one is grateful, and it seems rather bizarre if the computer would give the impression that it is grateful to the user, but it may be useful to indicate that a certain point is reached in the interaction. In Bunt (1999) we have described a mechanism of 'interactive pressures (IPs)' and 'reactive pressures (RPs)' responsible for the occurrence of SOM acts, where IPs arise under certain contextual conditions in the dialogue and give rise to initiative SOM acts, and RPs are created by initiative SOM acts and give rise to their reactive counterparts.

A similar analysis can be applied to other SOM acts. For instance, the IP to greet someone may arise when the contextual conditions are such that perceptual contact is possible. Therefore, an utterance like *"Good morning"* carries the information that the sender is perceptually 'present'. (Incidentally, this may also explain why an utterance like *"Hello* can be used both to assess contact and to perform a greeting.)

These examples illustrate that a context-change analysis reveals that SOM acts, often thought to be important 'just' for politeness, do in fact add information to the current dialogue context that is relevant for successful communication, and also show that differences in internal structure correspond to crucial differences in the information they convey.

Another interesting finding in the experiment that is worth mentioning is that there are great inter-subject differences. Some subjects do a lot of thanking and greeting, and provide ample positive feedback; others do this very little or not at all. Indeed, subjects sometimes have a certain personal *style* of interacting, which shows up in their use of these elements. This suggests that the use of these acts is a matter of stylistic or strategic choice.

The context-change interpretation of dialogue acts provides for most types of act a clear rule-based mechanism for when they are appropriate, given a sufficiently detailed context model. For positive feedback acts, anticipatory IM acts and SOM acts the contextual conditions for their use are not so clear, and apparently judged differently by different speakers who use different styles or strategies. This suggests that a conversational dialogue system should use positive FB acts, anticipatory IM acts and SOM acts according to a certain strategy. Possible strategies should correspond to strategies that human dialogue participants apply, and should therefore probably best be learned by trying to discover human dialogue strategies in appropriately annotated dialogue corpora.

References

Bilange, E. (1991). *Modélisation du dialogue oral personne–machine par une approche structurelle: théorie et réalisation.* Ph.D. thesis, Univ. Rennes 1.

Bunt, H. (1989). Information dialogues as communicative action in relation to partner modeling and information processing. In M. Taylor, F. Néel, and D. Bouwhuis (eds.), *The structure of multimodal dialogue.* Amsterdam: North-Holland Elsevier.

Bunt, H. (1994). Context and dialogue control. *THINK Quarterly 3*(1), 19–31.

Bunt, H. (1996). Interaction management functions and context representation requirements. In *Proc. Twente Workshop on Dialogue Management in Natural Language Systems.* Enschede: Universiteit Twente.

Bunt, H. (1999). Dynamic interpretation and dialogue theory. In M. Taylor, F. Néel, and D. Bouwhuis (eds.) *The structure of multimodal dialogue II*, Amsterdam: Benjamins.

Bunt, H. (2000). Dynamic pragmatics and context specification. In H. Bunt and W. Black (eds.) *Abduction, Belief and Context in Dialogue; Studies in Computational Pragmatics*, Amsterdam: Benjamins.

Hulstijn, J. (2000). *Dialogue models for inquiry and transaction.* Ph.D dissertation, University of Twente, Enschede.

Moeschler, J. (1989). *Modélisation du dialogue, réprésentation de l'inférence argumentative.* Paris: Hermès.

Reeves, B. and C. Nass (1996). *The Media Equation.* Stanford: CSLI Publications.