

GRACEFUL INTERACTION IN MAN-MACHINE COMMUNICATION

Phil Hayes and Raj Reddy,

Computer Science Department,
Carnegie-Mellon University,
Pittsburgh, PA 15213.

Compared to humans, current natural language dialogue systems often behave in a rigid and fragile manner when their conversations deviate from a narrowly conceived mainstream, e.g. when faced with ungrammatical, unclear, or unrecognizable input, ambiguous descriptions, or requests for clarification of their own output. We believe that the time is now ripe to construct systems which can interact gracefully with their users when such contingencies arise. Graceful interaction is not a single skill, but a combination of several diverse abilities. We list these components, and describe one of them - the ability to communicate robustly. Detailed descriptions of all the components appear in [4], along with details of a system architecture for their integrated implementation.

1. Introduction

A great deal of interest has recently been shown in computer systems capable of engaging in a dialogue with a human being in more or less natural language. This interest is embodied in numerous systems including GUS [1], LIFER [5], PAL [6], PARRY [7], and PLANES [8], and in the work of Codd [2], Grosz [3], and others. Such systems typically respond accurately and appropriately to straightforward requests and questions or otherwise uneventful dialogue within their domain of discourse. Compared to human beings, however, the performance of current dialogue systems appears quite rigid and fragile. Most current systems cannot, for instance: respond reasonably to input not conforming to a rigid grammar; ask for and understand clarification if their user's input is unclear; offer clarification of their own output if the user asks for it; or interact to resolve ambiguous descriptions. A dialogue system cannot hope to interact naturally and gracefully with its users, unless it can cope with these and the many other contingencies, so common in ordinary human conversation.

Graceful interaction, then, involves dealing appropriately with anything a user happens to say, rather than just those inputs in the mainstream of a conversation. Even though little attention has been paid to it in work to date (PARRY [7] being the main exception), we believe that graceful interaction is of critical importance in making natural language interfaces (for both typed and spoken input) a practical reality, and thus in making computer systems more accessible to casual or naive users, and more pleasant and natural for everyone.

Graceful interaction is an idea whose time has come and is ripe for implementation. We believe that graceful interaction is not a monolithic skill, but can be decomposed into a set of abilities and behaviours, none of which is significantly beyond the current state of the art. These components of graceful interaction are listed in the next section. In this short note, it is only possible to give a few details on just one of the components; we discuss all of them fully in [4]. In that paper, we also: claim that besides being necessary, our list of components is sufficient for graceful interaction, at least in a certain

(large) class of application domains; propose an architecture for the integrated implementation of all the components mentioned; and give a worked example of our design in operation.

2. Components of Graceful Interaction

Graceful interaction is not a single monolithic skill. Rather, it seems to be composed of a number of diverse abilities and behaviours, including:

Flexible parsing: The ability to deal with naturally used natural language, with all the ellipses, idioms, grammatical errors, and fragmentary utterances it can contain.

Robust communication: The set of strategies needed to ensure that a listener receives a speaker's utterance, and interprets it correctly.

Focus mechanisms: The ability to keep track of what the conversation is about, as the items under discussion shift; this is important for the resolution of ellipsis and anaphora, as well as for continuity in the conversation.

Explanation facility: The ability to explain what it can and cannot do, what it has done, what it is trying to do, and why, both for response to direct questions, and as a fall-back when communication breaks down.

Identification from descriptions: The ability to recognize an object from a description; including the ability to pursue a clarifying dialogue if the original description is unclear.

While none of these components of graceful interaction has been entirely neglected in the literature, no single current system comes close to having most of the abilities and behaviours we describe, and several are not possessed by any current systems.

These components are based on phenomena observable in naturally occurring human dialogues. We believe it is very important for a gracefully interacting system to conduct a dialogue in as human-like a way as possible; if the strategies the system employs for clarifying its incomprehensions of the user, resolving ambiguous

descriptions supplied by the user, etc. are not the same as those a human would use in the same situation, then the user will feel that the interaction is not natural, and hence not graceful. Furthermore, most of the components involve *cooperation* between speaker and listener; if the user tries to employ any of these techniques, an inability of the system to cooperate will appear most ungraceful.

Unfortunately, the aim of being as human-like as possible must be tempered by the limited potential for comprehension of any foreseeable computer system. Until a solution is found to the problems of organizing and using the range of world knowledge possessed by a human, practical systems will only be able to comprehend a small amount of input, typically within a specific domain of expertise. Graceful interaction must, therefore, supplement its simulation of human conversational ability with strategies to deal naturally and gracefully with input that is not fully understood, and to steer a conversation back to the system's home ground.

Space restrictions make it impossible to discuss all the components of graceful interaction mentioned above (see instead [4]), so we will concentrate on just one of them - the set of techniques for robust communication.

3. Robust Communication and Implicit Confirmation

During the course of a conversation, it is not uncommon for people to misunderstand or fail to understand each other. Such failures in communication do not usually cause the conversation to break down; rather, the participants are able to resolve the difficulty, usually by a short clarifying sub-dialogue, and continue with the conversation from where they left off. Current computer systems are unable to take part in such clarifying dialogues, or resolve communication difficulties in any other way. As a result, when such difficulties occur, and they can be neither eliminated nor anticipated, a computer dialogue system is unable to keep up its end of the conversation, and a complete breakdown is likely to result; this fragility lies in stark and unfavourable contrast to the robustness of human dialogue.

Since a speaker typically has no way of telling whether his listener has received his message correctly, detection of communication difficulties must rest on some convention of acknowledgement, commonly agreed upon by speaker and listener. Explicit acknowledgement by a listener of everything a speaker says would be tedious for humans. Instead, people use a convention that we will call the principle of implicit confirmation:

A speaker assumes his message has been received and interpreted correctly by his listener, unless the listener indicates otherwise.

While this approach is obviously very efficient when there is no difficulty with communication, it places a large burden on the listener. Unless the listener can determine that he has not received the message correctly, the error will go

undetected, and the conversation may become quite confused. However, a human listener normally has enough expectations about the sorts of things the speaker might say to make this a rare occurrence.

Although we have been using the terms, "speaker" and "listener", the same communication difficulties can arise for typed dialogues. While the errors typically arise from different sources (e.g. misrecognition of words v. spelling errors), for present purposes, we can continue to blur the distinction between spoken and typed dialogues. In what follows, we consider two different methods by which a listener can indicate incomprehension, discussing how much, if at all, current systems use these techniques, and noting modifications required for systems with limited ability to comprehend. Examples will use dialogues between a hypothetical directory assistance system (S) and its user (U).

3.1. Explicit Indications of Incomprehension

The principle of implicit confirmation requires a listener to give an explicit indication whenever he fails to comprehend a message or suspects that he may have received a message incorrectly. In this section we deal with the most straightforward way a listener can do this: by asking the speaker what he said or meant. Such questions vary according to how much of the original utterance the listener thinks he understood, how informative he tries to be about the nature of the problem, and how much he wants to influence the speaker's subsequent reply.

Phrases such as "I beg your pardon" are the least informative way of indicating incomprehension. They provide the original speaker with no clue about what went wrong with the communication, and so are most appropriate when the failure is due to a transient problem such as a sudden noise. A speaker is much more likely to modify his communication in a way that will allow it to succeed at a second attempt if his listener gives him some information about the nature of the problem.

The clues the listener can provide depend on the degree to which he failed to understand the utterance. If he understood nothing at all, the only clue he can provide is what he expected the speaker to say:

S: This is directory assistance.

U: GARBLE

S: Excuse me? May I find you a number?

If he has understood something, he can show what he has understood while indicating incomplete comprehension:

U: What is the number for GARBLE?

S: Whose number did you want?

If he has understood enough to narrow the possibilities down to a small number, he can list them. The information given by these various strategies to the original speaker allows him to concentrate on imparting the information that did not get across.

The way in which incomprehension is indicated can influence the form of the subsequent clarification. As Codd

[2] points out, a gracefully interacting system with limited power of comprehension should therefore phrase its indications of incomprehension to maximize its chances of an understandable reply. For instance, if a system does not understand "extension" in "What is the extension for Jim Smith?", it is unwise to reply "What is an extension?", but better to ask "Do you want Jim Smith's number or his address?". On the other hand, such directiveness is never guaranteed success, and in particular, a user cannot always be expected to cooperate by choosing one of the alternatives in a multiple choice question. For instance, reasonable responses to the last question include "The number", "The first one", "Both", "I want to phone him", "Do you know where he is now?", "I want the number, but I meant Joe Smith". This raises the possibility of a user's clarification also being misunderstood, and of a sequence of misunderstandings and clarifications failing to converge. In such cases, a gracefully interacting system will have to become more and more explicit about the nature of its problems and its limitations, and rely on the user to accommodate himself.

In summary, a gracefully interacting system must be able to express its own incomprehension of or uncertainty about its user's input, and be able to deal with the user's reports of his own problems in comprehending the system. While most systems can indicate incomprehension, most do it in a more or less uninformative and undirective way. As far as we know, no current system can respond appropriately to explicit complaints of incomprehension by its user.

3.2. Echoing

Echoing is a way of confirming uncertain interpretations without the requirement for a reply associated with an explicit request for confirmation. If a listener wants to be sure that his interpretation of the speaker's utterance (or more commonly part of it) is correct, he need only echo his interpretation. If the original speaker does not comment on the echo, it is implicitly confirmed. Thus:

U: What is the number for ?ter Smith?
 S: Walter Smith ... His number is 5592.
 U: Thank you.

Of course, the original speaker still has the option of confirming the echo explicitly; and if the echo is incorrect, he must indicate that explicitly.

A gracefully interacting system should be able to issue echoes when it is uncertain in its comprehension, and to accept any corrections its user offers. It must also be constantly on the watch for echoes by the user of what it says. As far as we know, none of these aspects of echoing has ever been implemented in a dialogue system.

4. Summary

In this short note, we have tried to show the importance of graceful interaction for man-machine communication, outline the scope of the problems involved in constructing gracefully interacting systems, and provide a small amount of detail about one of those problems. We believe that

graceful interaction is an idea whose time has come, that it can be decomposed into a set of relatively independent components, none significantly beyond the current state of the art, and that these components can be made to perform together in a truly gracefully interacting system. Details of the proposed components, including robust communication, flexible parsing, focus mechanisms, explanation facility, identification from descriptions, and generation of descriptions, can be found in [4], along with a description of the system architecture in which we propose to integrate all these components. A gracefully interacting system conforming to this architecture is currently under implementation.

References

1. Bobrow, D. G., Kaplan, R. M., Kay, M., Norman D. A., Thompson, H., and Winograd, T. GUS: a Frame-Driven Dialogue System. *Artificial Intelligence* 8 (1977), 155-173.
2. Codd, E. F. Seven Steps to RENDEZVOUS with the Casual User. In Klimbie, J. W. and Koffeman, K. L., Ed., *Proc. IFIP TC-2 Working Conf. on Data Base Management Systems*, North Holland, Amsterdam, 1974, pp. 179-200.
3. Grosz, B. J. The Representation and Use of Focus in a System for Understanding Dialogues. *Proc. Fifth Int. Jt. Conf. on Artificial Intelligence*, MIT, 1977, pp. 67-76.
4. Hayes, P. J., and Reddy, R. An Anatomy of Graceful Interaction in Man-Machine Communication. Tech. report, Computer Science Department, Carnegie-Mellon University, 1979.
5. Hendrix, G. G. Human Engineering for Applied Natural Language Processing. *Proc. Fifth Int. Jt. Conf. on Artificial Intelligence*, MIT, 1977, pp. 183-191.
6. Sidner, C. A Progress Report on the Discourse and Reference Components of PAL. A. I. Memo. 468, MIT A. I. Lab., 1978.
7. Parkison, R. C., Colby, K. M., and Faught, W. S. Conversational Language Comprehension Using Integrated Pattern-Matching and Parsing. *Artificial Intelligence* 9 (1977), 111-134.
8. Waltz, D. L. An English Language Question Answering System for a Large Relational Data Base. *Comm. ACM* 21, 7 (1978), 526-539.