THE PROVISION OF CORRECTIVE FEEDBACK IN A SPOKEN DIALOGUE CALL SYSTEM

Sarah Davies and Massimo Poesio

Human Communication Research Centre, University of Edinburgh, EH8 9LW, UK.
{sarahd, poesio} @cogsci.ed.ac.uk

ABSTRACT

In this paper we report on the development of a spoken dialogue system for computer aided language learning (CALL), and explore some of the issues involved in the incorporation of a corrective feedback module. We initially developed a small prototype system, and tested it for usability with visiting students of English as a foreign language. In the light of the positive results we obtained for this, we began to develop a more advanced system, with the aim of investigating how spoken dialogue systems might best be tailored to help language learning. The issue we focussed on was the kind of feedback on errors which might be most useful to the learner. We show the types of feedback we have considered, and highlight some of the problems associated with providing different types of feedback.

1. INTRODUCTION

The long-term aim of our work is to find out whether spoken dialogue systems can be used to help people learn languages. In order to answer this question, two issues must be resolved: that of usability (whether the systems are robust, the students are able to negotiate the interface and complete the task, whether they enjoy it and want to use the system again) and that of usefulness (do these systems actually help people to learn and improve long-term language development?). In order to address the first issue, we developed two prototype systems, and tested them with foreign learners of English (section 2), with promising results. We then moved to the issue of usefulness. Rather than designing and developing a system on the basis of our beliefs and preconceptions about language learning, and then testing this, we decided to approach this very large problem incrementally, by tackling one aspect of an educational system at a time. We are currently looking at the provision of corrective feedback when the system detects that the learner has made an error. We are developing a system with a number of different feedback strategies, so that these strategies can be tested with students to see whether any appear to favour the their language development. Pedagogical issues in corrective feedback are outlined in section 3, and we then describe some of the possible feedback strategies which could be employed by a system (section 4) and look at some preliminary issues of implementation (section 5).

2. A PRELIMINARY ASSESSMENT OF USABILITY

In order to make a preliminary assessment of the feasibility of using simple spoken dialogue systems in language learning, we used the CLSU Rapid Application Developer (RAD) [1] to develop prototype systems for use in teaching English as a foreign language. These were then tested with small groups of visiting students.

2.1. A Hike in the Highlands

After testing the performance of the RAD's speech recogniser on a variety of tasks, we built and tested a simple audio-only prototype system. This was designed to practise pronunciation within a game-like format, in which the students affect their progress through the scenario by choosing what they want to do at each stage from a list of options. As students were able to use this without many problems, and the performance of the recogniser was very good, we felt that the complexity of the task could be increased if the interface was also improved. A further prototype, A Hike in the Highlands, was therefore developed [2].

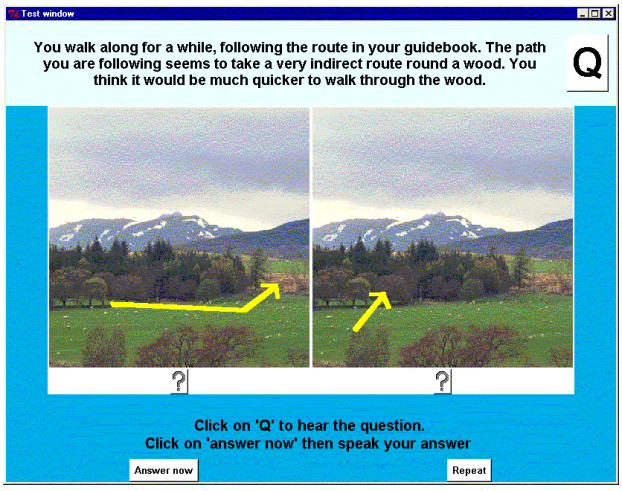

Figure 1. Screen from A Hike in the Highlands.

A Hike in the Highlands aimed to provide more general production practice for the student, with the focus on being able to put an appropriate phrase together, rather than on very accurate pronunciation. The aim in this scenario was to go on a hike in the Highlands and make it back safely after an enjoyable day out rather than being left, cold and hungry, on the mountain. In order to complete their task, the students had to choose what to do at a number of points. The system employed a graphical interface, which meant that the options available to the student at each stage could be presented through images. We felt that this was more encouraging of creative language production than the use of a spoken or written list of options,

which would allow the student to simply echo the language used by the system. (In the event of difficulty, students can hear a suggested phrase for each image.) The buttons give the user control over the pace of the dialogue and the number of repetitions of the prompt. The multimedia interface also means that the student can use clues from the text and images to support their understanding of the spoken audio, which can be problematic for many learners.

The recognition strategy used is multiple word-spotting: words or phrases are identified in various parts of the user's utterance, without trying to recognise every word. For example, if the target phrase was something like *Take the shortcut through the wood*, one of the combinations we might search for would be *take*, *shortcut*, *through* and *wood*, allowing any other words to come between these.

2.2. Testing

The Hike system was tested by 11 Spanish and Italian EFL students, alone and in pairs. They were talked through the first screen and then left to complete the rest of the game, though someone was on hand to answer any queries.

The students used the system successfully and were able to understand the scene and choose an option at each stage, even though some of them found the language content very challenging. In a questionnaire filled in after using the system, all 11 students said they found the system *easy to use*, 6 found the pictures *very helpful* and 5 *quite helpful*, and 7 thought the system would be *quite helpful* and 4 *very helpful* in learning English.

The performance of the recogniser on this less constrained task was considerably worse than it had been on the simpler system. The 9% rate of non-recognition (the failure to recognise the student's utterance with sufficient confidence on the first attempt) was not too problematic, as the students were able to repeat their answer and continue. More worrying was the high rate of misrecognition (the erroneous recognition of a student utterance as something other than what the student said, without the system being aware that it has made a mistake). Although no explicit feedback is provided by this system, the misinterpretation of the student's utterance by the system implicitly gives incorrect feedback, making the student feel that she has not communicated effectively. It is therefore clear that the parameters of the recogniser need to be adjusted to minimise this kind of mistake.

3. CORRECTIVE FEEDBACK IN LANGUAGE LEARNING

As mentioned in the introduction, our long-term aim is to investigate whether spoken dialogue systems help people learn languages, but as this is such a large question, we are currently looking at just one sub-issue, that of corrective feedback.

Spoken dialogue systems offer the potential for a student to practise speaking the target language without embarrassment and without needing the presence of a speaker of the target language. However, this functionality alone may not necessarily help learning, as it has been suggested [3, 4, 5] that the benefit of output, or practice, in language learning stems at least in part from its tendency to elicit feedback on language use. More generally, the benefit or otherwise of various types of feedback is still a matter of some controversy in second language acquisition research [6]. We hoped to develop a system which would enable us to investigate this question.

Human feedback to the learner on her language can take many forms, from an explanation of the grammar to blank incomprehension or even laughter [7]. In a human-machine dialogue, it can also be given in different presentational formats, which may interact with students' own language learning strategies in different ways [8]. Our first goal was to look not at the presentation but the content of the feedback, varying the type and amount of information given to the student.

As the initial step towards investigating the effectiveness of different feedback strategies, we needed to build a system which could be set to give one (or give predominantly one) of a number of different possible types of feedback, while holding all other factors constant. We then hope to have students use the system over a term, and investigate whether there is any difference in language improvement which can be correlated with the use of a certain feedback strategy.

3.1 Types of Corrective Feedback

The corrective feedback strategies we intend to test are taken from Lyster and Ranta's 1997 study of classroom feedback strategies [9]. They observed student-teacher interactions in French immersion classrooms in Montreal, and found that the corrective feedback given to the students could be classified into six types:

- Explicit correction: the explicit provision of the correct word or part phrase, usually making clear that this is a correction - e.g. *you mean..., you should say...*
- Recast: the teacher's reformulation of all of part of the student utterance, minus the error, without making it clear that this is a correction.
- Clarification request: *What? What do you mean?* (only coded in response to language error).
- Metalinguistic feedback: comments, information or questions regarding the well-formedness of the student's utterance, but without giving the correct form: *that's not quite right, is that right?, 'problema' is masculine*.
- Elicitation: getting the student to give the correct form by pausing for her to continue the sentence, or by asking the student to reformulate the utterance.
- Repetition: the repetition, in isolation, of the student's utterance, usually with error intonationally marked. (This was found to co-occur with all other types of feedback.)

In addition, 38% of errors received no correction.

Lyster and Ranta compared these strategies by looking at what happened in the turns immediately following the provision of feedback: did the student try to correct her previous utterance?, was that reformulation correct?, and was the correct form initially suggested by the student concerned, or by the teacher or another student? They found that recast, the most commonly used form of feedback, had the lowest rate of uptake: students tried to correct their previous utterance in 31% of cases, with about half of the reformulations being correct, even though the correct form had just been provided by the teacher. Explicit correction had a good rate of correct student reformulations, but, due to the form of feedback, the correct forms were always generated by the teacher, rather than the student. The use of elicitation always required the student to attempt to generate the correct form themselves, and as such produced the highest rate of correct student-generated repairs. Although metalinguistic feedback had a lower rate of uptake overall, a similar proportion of the attempted reformulations were correct student-generated repairs.

As mentioned in the previous section, we plan to investigate whether these differences in the patterns of student behaviour following feedback in the classroom would result in any difference in language learning over a 12-week term if the students were exposed to one kind of feedback from a regularly used spoken dialogue CALL system.

4. THE INTEGRATION OF CORRECTIVE FEEDBACK IN A SPOKEN DIALOGUE SYSTEM

Our experimental aims dictate that we build a spoken dialogue system which is able to offer a range of corrective feedback strategies, and can be set to provide only one of these, while remaining the same in all other ways. We therefore had to consider how the human-human strategies outlined by Lyster and Ranta might be implemented in a human-machine dialogue. We described the six strategies by their component actions (Table 1), and then analysed them in terms of the information needed by a CALL system's recogniser and dialogue manager in order to provide the relevant feedback to the student (Table 2).

In addition to the characteristic actions given in Table 1, the strategies also differ as to what the teacher or system does immediately afterwards: whether or not a correction turn from the student is awaited before moving on with the dialogue. Elicitation, for example, involves a pause until the student attempts a repair, whereas classroom teachers may often move straight on after recast. These features could be implemented on a per-strategy basis in a dialogue system, but it might be deemed more user-friendly always to give the student the option of not responding to the correction and continuing with the conversation.

It can be seen from Table 2 that certain strategies require very different information from others. This can be illustrated by a consideration of the two strategies which we have currently implemented, metalinguistic feedback and recast. In order to provide metalinguistic feedback, the system has to be able to identify the grammatical or lexical misconception or slip which is causing the error, and point to the location of the error in the student's version. For recast neither of these are directly required, as all that is needed is the production of a correct version of the learner's phrase.

FEEDBACK STRATEGY	ACTIONS				
	Indicate there was an error	Signal problem with mean-ing	Indicate where error was	Give correct form	Explain error
Explicit correction	YES	no	YES	YES	no
Recast	no	no	no	YES	no
Clarification request	no	YES	no	no	no
Metaling. feedback	YES	no	YES	no	YES
Elicitation	YES	no	optional	no	no
Repetition	YES	no	YES	no	no

Table 1. The description of six feedback strategies in terms of the actions performed by the teacher or dialogue system.

FEEDBACK STRATEGY	SYSTEM NEEDS TO KNOW				
	Correct full phrase	Exact words of learner	Locat-ion of error	Correct version of error	Gramm-ar causing error
Explicit correction	no	no	YES	YES	no
Recast	YES	no	no	no	no
Clarification request	no	no	no	no	no
Metaling. feedback	no	no	YES	no	YES
Elicitation	no	no	no	no	no
Repetition	no	YES	YES	no	no

Table 2. The information needed by the recogniser and dialogue manager of a spoken dialogue system in order to provide the different types of feedback.

As the system already needs to know the possible correct utterances at each stage, the most straightforward way to provide recast would be to pattern match the user's incorrect utterance against these. However, this approach does not help us to provide metalinguistic or other types of feedback. An approach we are exploring for metalinguistic feedback is predicting the likely user errors on the basis of teachers' intuitions, and then linking each error with one of a number of broad types of grammatical error (currently we have 13 broad types of error for student mistakes with present-tense questions and negatives in English). Feedback is then given on the basis of the broad type of error encountered. If we do wish to adopt an integrated approach to error correction, the recast could be generated from the information we have already gathered about the user error, if we also encode detailed knowledge for the system about how this error is to be corrected. (The metalinguistic feedback given to the student does not specify exactly how to correct the error, as it is intended that the student is required to generate the correct version herself). We have yet to establish whether this more integrated approach is any more than theoretically pleasing, as it may prove to be the case that different types of feedback are best provided through different mechanisms.

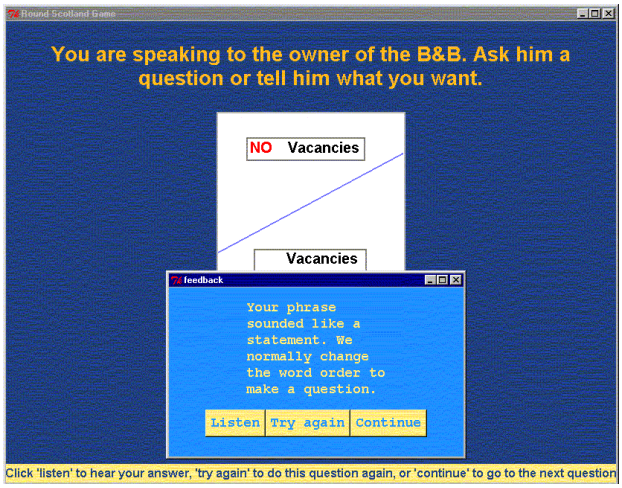

Figure 2. This screen shows an example of how metalinguistic feedback is presented to the user within our new prototype. If the same error is made again, more detailed feedback is provided.

4.1 Implications for the performance of the speech recogniser

It is clear that, in order to provide the kind of feedback we have just discussed, the recognition strategy used needs to be more accurate than that outlined in Section 2.1. Knowing that someone has said *rooms, available* and *tonight* in the context of the opening question to the owner of a hotel may be enough to allow an informed guess as to what the intention of the question was, but it is certainly not enough to be able to tell whether it was correctly formed. We have therefore implemented a new recognition strategy, in which the recogniser attempts to identify nearly every word spoken by the user (generally articles are excepted due to their very low phonological salience). When only possible correct utterances are considered, the recogniser performance is no difference from that seen with the Hike system, but if it is asked to decide whether a given utterance is any of a few correct utterances and a great many predicted erroneous utterances, performance drops sharply.

We are currently investigating how the performance might be improved. However, it is very unlikely that misrecognitions will disappear altogether, so when an grammatical error is detected, we provide a facility for the student to listen to her utterance again and decide whether the error identified did in fact occur. It may also be useful in this context to make the output of the recogniser available to the student directly.

5. CONCLUSION

Our preliminary investigations with the Hike in the Highlands prototype suggest that it is feasible to build usable spoken dialogue systems for CALL with currently available technology. We are now starting to investigate whether the provision of various types of corrective feedback within a dialogue system can aid language learning, and as a first step towards this are building a new prototype incorporating a number of different corrective feedback strategies.

6. REFERENCES

1. Sutton, S. and Kaiser, E. *The CSLU Rapid Prototyper: Version 1.8*, CSLU, OGI, 1996
2. Davies, S. and Poesio, M. "A CSLUrp-Based Spoken Dialogue System for TEFL,". *Proc. STiLL*, 1998.
3. Levelt, W. J. M. *Speaking: from Intention to Articulation*. MIT Press, Cambridge, MA, 1989.
4. de Bot, K. "A Bilingual Production Model: Levelt's 'Speaking' Model Adapted," *Applied Linguistics*, 13, 1992.
5. de Bot, K. "The Psycholinguistics of the Output Hypothesis," *Language Learning*, 46, 1996
6. Schachter, J. "Corrective Feedback in Historical Perspective," *Second Language Research*, 7(2), 1991.
7. Schachter, J. "Three Approaches to the Study of Input," *Language Learning*, 36(2), 1986.
8. Eskenazi, M. and Hansma, S. "The Fluency Pronunciation Trainer," *Proc. STiLL*, 1998.
9. Lyster, R. and Ranta, L. "Corrective Feedback and Learner Uptake," *Studies in Second Language Acquisition*, 18, 1997.