A Study of Feedback Strategies in Foreign Language Classrooms and Tutorials with Implications for Intelligent Computer-Assisted Language Learning Systems

Anita Ferreira, *Department of Spanish Linguistics, Universidad de Concepción, Concepción, Chile*
aferreir@udec.cl

Johanna D. Moore, *Division of Informatics, University of Edinburgh, Scotland, UK*
J.Moore@ed.ac.uk

Chris Mellish, *Department of Computing Science, University of Aberdeen, Scotland, UK*
cmellish@csd.abdn.ac.uk

Abstract. This paper presents two new corpus-based studies of feedback in the domain of teaching Spanish as a foreign language, concentrating on the type and frequency of different feedback moves, as well as what happens in the moves that follow the feedback. In particular, as well as looking at positive feedback, it concentrates on two general kinds of negative feedback strategies: (1) Giving-Answer Strategies (GAS), where the teacher directly gives the desired target form or indicates the location of the error, and (2) Prompting-Answer Strategies (PAS), where the teacher pushes the student less directly to notice and repair their own error. Investigating the GAS/PAS distinction sheds light on the relative importance for Intelligent Computer-Assisted Language Learning (ICALL) systems of knowledge construction from interaction, which many believe is crucial for effective learning from ITS. The main finding here is that, although GAS occur more frequently than PAS in both corpora, it is the PAS that are more effective, in terms of eliciting explicit repairs by the students.

The first study takes place in a classroom context, whereas the second, smaller, study looks at tutorial interactions. This makes it possible to investigate the extent to which the mode of interaction influences the frequency and effectiveness of feedback moves, as well as to look at how concepts such as "wait time" are relevant to explain moves that are ineffective.

The paper concludes by using these results to make recommendations about how to choose appropriate feedback moves in ICALL systems.

Keywords. Intelligent tutoring systems for foreign language, intelligent computer assisted language learning, feedback strategies in foreign language, technological applications in applied linguistics

INTRODUCTION

Feedback and guidance moves (such as prompting, hinting, scaffolding, and pumping) have been investigated at some length by researchers working on Intelligent Tutoring Systems (ITS) for teaching procedural skills in domains such as algebra, geometry, physics and computer programming (Anderson et al., 1995; Merrill et al., 1992; Hume et al., 1996; Chi et al., 2001). However, little attention has been paid to feedback in the domain of ITS for foreign languages. This situation seems to have arisen for at least the following reasons:

- *The specific and complex nature of errors and corrective feedback in ITS for Foreign Language.*

  Errors and corrective feedback are a natural part of language learning. **Errors** can be defined as deviations from the norms of the target language (Ellis, 1997). **Corrective Feedback** is an indication to a learner that his or her use of the target language is incorrect. According to Ellis (Ellis, 1997), the main finding of studies of error treatment is that it is an enormously complex process. This can be seen in the different models of feedback (Long, 1977; Chaudron, 1977; Lyster & Ranta, 1997) and the different taxonomies of feedback strategies (Allwright, 1975; Chaudron, 1977; Seedhouse, 1997; Lyster & Ranta, 1997) that have been proposed. Some of these strategies are markedly different from those typically found in ITSs for procedural domains.

- *The relative merits of different types of feedback are still not fully understood in the area of Second Language Acquisition (SLA).*

  SLA research on feedback reveals that teachers have a wide variety of strategies available for the treatment of students' errors (Allwright, 1975; Chaudron, 1977; Seedhouse, 1997; Lyster & Ranta, 1997). However, it is only recently that systematic studies into the type, frequency and effectiveness of different feedback strategies have been carried out (Doughty & Varela, 1998; Seedhouse, 1997; Lyster, 1998b; Lyster, 1998a). The results of these empirical investigations indicate that (1) the relative merits of different types of feedback are not fully understood, and (2) the relative effectiveness of feedback strategies depends on multiple variables, including the particular aspects of the language being corrected, conditions relating to the provision of teacher correction, and characteristics of the students (e.g. sophisticated grammatical explanations are not appropriate for beginning students).

  Thus, we do not yet have the detailed knowledge needed to implement systems that could use the most appropriate and effective strategy in a given situation. For this reason, it is important to carry out further research aimed at providing empirical evidence to inform the design of such systems.

- *The lack of empirical research on the effectiveness of ICALL for foreign language, in general, and feedback strategies in particular.*

  ICALL and ITS systems for foreign language typically focus on the use of parsing techniques from natural language processing (NLP) to analyze students' responses and identify errors or missing items. These NLP capabilities allow systems to handle sophisticated types of feedback strategies, such as meta-linguistic feedback[1] and "error reports"[2] to correct particular student errors based on a textual analysis (Criswell et al., 1991; Sams, 1995; Levin & Evans, 1995; Nagata, 1995; Nagata, 1997). However, the unnatural and obstructive form that this feedback takes is only suitable in the context of grammar exercises, where there is a focus on linguistic form, and for advanced learners who are capable of handling a series of involved technical explanations. As a consequence, ITS and ICALL systems have not yet incorporated the strategies that are typically used by second and foreign language teachers and studied by SLA researchers (Allwright, 1975; Chaudron, 1977; Seedhouse, 1997; Long, 1977; Lyster & Ranta, 1997; Lyster, 1998a; Doughty & Varela, 1998).

[1] Comments, information or questions related to the well-formedness of the student's utterance.

[2] These inform the student that an error was made and give the location of the error in the sentence.

- *The lack of awareness of results from SLA research by researchers in the ICALL and ITS communities.*

  Most ITS and ICALL systems appear to be created without reference to the many research studies concerning language learners' abilities, that is, how they may best learn languages, and how teachers deal with students' errors (Bull, 1997). We believe that to improve the effectiveness of ICALL systems for teaching foreign language, designers should pay more attention to the results emerging from SLA research. In particular, we should not assume that we can devise governing tutorial principles to serve as control structures for a system without having examined, in detail, natural classroom and tutorial interactions between teachers and learners (Tomlin, 1995). Nor should we assume that all teaching strategies observed in human student-tutor interactions are suitable for implementation in ICALL systems for FL.

Despite these problems, the value of the feedback component is unquestioned in the area of ICALL for foreign language (Garrett, 1995). However, it is necessary to determine the extent to which feedback actually contributes to the language development of the learner, and therefore it is crucial that we obtain more detail about how learners respond to feedback strategies, and whether the strategies result in a change in their learning. To this end, we have explored empirical evidence about the type, frequency and effectiveness of feedback strategies based on studies involving three different learning contexts: an observational study of face-to-face classroom interactions, a case study of one-on-one tutorial interactions, and an experimental study in which students interacted with a web-based tutoring program.

In this paper, we focus on the first two studies of feedback strategies in foreign language: the main study of classroom interactions and a smaller study of tutorial interactions, comparing feedback strategies in the two teaching modes. We propose here that the incorporation of effective feedback strategies into ICALL systems for Spanish as a foreign language can be informed by (1) the analysis of feedback strategies used in Spanish as FL classes/tutorials, and (2) SLA research findings on the effectiveness of corrective feedback.

POSITIVE AND CORRECTIVE FEEDBACK STRATEGIES

In our studies, we investigated feedback strategies for both positive and corrective feedback. Regarding positive feedback, we are interested in determining what the teacher does after the student provides a correct answer, other than provide **acknowledgement** or **acceptance** of the correct answer. We thus considered two additional types of **positive feedback** which have been discussed in the SLA literature (Ellis, 1997):

- **Repetition:** The teacher repeats the student's correct answer. For example, Student: "club de jóvenes" (*youth club*); Tutor: "club de jóvenes".
- **Rephrasing:** The teacher accepts the student's answer but aims to expand the student's knowledge, to polish the utterance structure, or to show a new structure which rephrases the answer given by the student using different words, and in some cases, adds new information. For example, S:

"sacerdote"; T: "sacerdote, padre, párroco... ok... son palabras culturales" (*priest, father, parish priest ... ok ... are cultural words*).

In second language teaching, **Corrective Feedback** is an indication to a learner that his or her use of the target language is incorrect. In our studies, we classified corrective feedback strategies identified in the SLA literature into two groups:

1. **Giving-Answer Strategies (GAS)**: Types of feedback moves in which the teacher directly gives the target form corresponding to the error in a student's answer, or shows the location of the student's error. These include:

   (a) **Repetition:** The teacher repeats the error or the portion of the learner's phrase containing the error, using stress or rising intonation to focus the student's attention on the problematic part of the utterance. For example, S: "Future" (Incorrect tense); T: "¿Future?"

   (b) **Recast:** Reformulation of all or part of the student's answer, providing the target form. For example, S: "En el segundo piso, hay cuatro dormitorio y dos baño." (*On the second floor, there are four bedroom and two bathroom.*) T: "Qué grande es tu casa! Tiene cuatro dormitorios y dos baños." (*What a big house you have. It has four bedrooms and two bathrooms.*)

   (c) **Explicit correction:** The teacher provides the correct target form. For example, S: "Cuando ella andó." *(When she went)*; T: "andaba." This differs from recast because the teacher directly corrects the error without rephrasing or reformulating the student's answer.

   (d) **Give answer:** Used in cases when the student does not know or is unsure of the answer. For example, S: "Ella compró mucha fruta y..." (Student can not finish his answer because he does not know how to say vegetables). T: "Fruta y verduras." (Teacher completes the answer with the word verduras.)

   Our definitions of the repetition and recast strategies are based on those used in Doughty and Varela's (Doughty & Varela, 1998) study.

2. **Prompting-Answer Strategies (PAS)**. Types of feedback moves in which the teacher pushes students to notice a language error in their response and to repair the error for themselves. We have called this group prompting answer strategies because of the similarity these strategies bear to the notion of "prompting" described in (Chi et al., 2001). This group includes three types of strategies:

   (a) **Meta-linguistic cues:** The teacher provides information or asks questions regarding the correctness of the student's utterance, without explicitly providing the target form. For example, S: "Compra" (to buy); T: "Tienes que poner un condicional." *(You have to use a conditional.)*

   (b) **Clarification requests:** Questions intended to indicate to the student that his/her answer has been misunderstood due to a student error, or that the utterance is ill-formed in some way and that a repetition or a reformulation is required. Clarification requests often include phrases such as "Pardon me.", "What?", "What do you mean?". For example, S: "Me gustaría un jugo de cilantro." (*I would like a coriander juice*); T: "¿Qué cosa?" (*What do you mean?*).

(c) **Elicitation:** The teacher encourages the student to give the correct form by pausing to allow the student to complete the teacher's utterance, by asking the student to reformulate the utterance, or by asking questions to elicit the correct answer, such as "How do we say that in Spanish?" For example, T:"¿Qué debe hacer Roberto?" (What does Roberto need to do?) S: "Brush his teeth". T: "How do we say that in Spanish?"

The prompting-answer strategies are based on a study by Lyster (Lyster, 1998a), which grouped four feedback strategies (elicitation, meta-linguistic cues, clarification requests, and repetition of error) under the rubric of **Negotiation of Form**. Lyster distinguished Negotiation of Form from recast and explicit correction because it provides learners with signals that facilitate peer- and self-repair, in a way that rephrasing or correction of student utterances do not.

There is scope for debate about where repetition should belong in this classification. Repetition is unlike other GAS, in that it does not include corrected material and goes further towards requiring an explicit response from the student. However, unlike Lyster, we do not include repetition of errors in our definition of prompting-answer strategies, because, by including elements of the student's answer, it presents features of recast. In fact, Lyster (Lyster, 1998b) distinguishes four types of recast, one of which is defined as "an isolated interrogative recast which seeks confirmation of the learner's message by correctly reformulating all or part of the utterance with rising intonation and no additional meaning". This would seem to allow repetition as a special case. Similarly, Doughty and Varela (Doughty & Varela, 1998) consider repetition of the error part of corrective recasting so as to draw attention with stress and rising intonation.

By distinguishing the PAS and GAS groups in our study, we hope to gain further insight into the relative merits of feedback strategies that encourage students to attempt to generate or construct the correct form themselves (PAS), versus those in which the teacher resolves the language error either by indicating the location of the error or providing the target form (GAS). This distinction is motivated by Chi et al.'s (Chi et al., 2001) study arguing for the benefits of constructive learning, in which the student as an active learner constructs an understanding by interpreting new material in the context of prior knowledge by, for example, making inferences, elaborating the material, or integrating material. Knowledge construction may occur as a result of self-explaining (either spontaneously (Chi et al., 1989) or as the result of elicitation (Chi et al., 1994)), asking questions, responding to the teacher's questions, etc.

As in many previous empirical studies in the ITS literature and building on a long tradition of empirical work on human tutoring, Chi et al. (Chi et al., 2001) studied one-on-one human tutorial interactions, in an attempt to identify the features that make tutoring such an effective learning intervention. Their studies suggest that one-on-one tutoring is effective because it provides students with opportunities to engage in constructive activities, and they argue that ITSs should implement ways to elicit constructive responses from students. However, Chi et al.'s studies were carried out in a complex conceptual domain (the human circulatory system), where they contrasted a condition in which tutors provided explicit feedback and didactic explanation, with a condition in which tutors were instructed to prompt students to construct their own answers. In our work, we aim to determine whether these results carry over to the domain of ICALL. To this end, we have grouped the feedback strategies observed by SLA researchers

into the giving- and prompting-answer categories in order to investigate the frequency with which foreign language teachers use the different types of strategies to correct each type of error, and the relative effectiveness of the strategies.

Further, because much of the work in SLA is based on empirical work in classrooms, another of our goals is to understand the similarities and differences between the strategies that teachers use in classrooms and those that are used in tutorial interactions. Chi et al. suggest that the interactive nature of one-on-one tutoring provides greater opportunities for students to engage in constructive activities, e.g. repairing their own errors, than they have in traditional classrooms in which there are 20 or more students. To shed more light on this issue, we compare the results from our study of tutorial interactions with the results obtained from our study of classroom interaction in terms of type, frequency and effectiveness of feedback strategies.

More precisely, we set out to answer the following research questions:

1. *What are the most frequent types of positive feedback used by teachers of Spanish as FL after a student's correct answer?*
2. *What are the most frequent types of corrective feedback used by teachers in Spanish as FL, and how do they relate to learner errors? In particular, what are the relative frequencies of GAS and PAS?*
3. *What are the most effective types of corrective feedback? In particular, are there differences in the apparent effectiveness of GAS and PAS?*
4. *Are there differences in type, frequency, and/or effectiveness of the strategies in the two teaching modes?*

The remainder of the paper reports two studies aimed at answering these questions. In the next section, we describe relevant prior research on feedback in ICALL and ITS for foreign language teaching. In the section on the analysis of classroom and tutorial interaction, we describe the data that we collected for the two studies and the analysis scheme we applied. The following section presents the results of two studies focused on determining the most frequent and effective feedback strategies in Spanish as FL learning, for classroom and tutorial teaching modes. There is then a section which presents further analysis of situations in which GAS and PAS seem to be ineffective in the studies. We then discuss these results, and present their implications for the design of ICALL systems for foreign language learning. Finally, we present our conclusions and ideas for future work.

RELATED WORK

In this section, we briefly review related work that sheds light on the nature of feedback and our research questions.

Corrective Feedback in Second Language Acquisition

There is a considerable literature dealing with error treatment in SLA. Much of this literature is devoted to addressing the questions of whether, when, and how errors should be corrected and who should cor-

rect them (Ellis, 1997). Apart from general instruction, the primary role of language teachers is often considered to be the provision of error correction, a form of corrective feedback, and positive sanctions or approval of learners' production.

Errors and corrective feedback constitute a natural part of the teaching-learning process in a foreign language. Errors can be defined as deviations from the norms of the target language (Ellis, 1997). They reveal the patterns of learners' development of inter-language systems, showing where they have over-generalized a foreign language rule or where they have inappropriately transferred a first language rule to the foreign language (Lightbown & Spada, 1999). Corrective feedback is an indication to a learner that his or her use of the target language is incorrect, and includes a variety of responses that a language learner receives. Corrective feedback can be explicit (e.g. "No you should say goes, not go.") or implicit (e.g. "Yes, he goes to school every day."), and may or may not include meta-linguistic information (e.g. "Don't forget to make the verb agree with the subject.") (Lightbown & Spada, 1999).

The effectiveness of corrective feedback depends on multiple factors, including: (1) the particular features of language being corrected, (2) the conditions relating to the provision of teacher correction, (3) the appropriateness of the student's stage in his/her language learning process to benefit from the correction (timing of the process), and (4) the ability of learners to notice a gap between what they want to say and what they can say, leading them to differentiate what they do not know from what they know only partially (the **noticing function**) (Schmidt & Frota, 1986). Moreover, corrective feedback is most beneficial when it occurs in response to naturally-occurring errors or in the context of ongoing efforts to communicate.

In cognitive terms, the function of corrective feedback is to provide information that learners can actively use in modifying their behavior. The information available in the feedback allows learners to confirm, disconfirm, and possibly modify the hypothetical, transitional rules of their developing grammars. However, these effects depend on the learner's developmental stage and ability to notice the information available in the feedback. Student-generated repairs may also be important for learning because they provide learners with opportunities to automate the retrieval of target-language knowledge. When repair is self-generated, learners draw on their own resources; that is, they actively confront errors in ways that may lead to revisions of their hypotheses about the target language (Chaudron, 1988).

Empirical Studies of Corrective Feedback

Research on corrective feedback reveals that teachers have a wide variety of techniques available for the treatment of student's errors. Corrective feedback in the form of recast has been the focus of a number of recent studies. Long et al. (Long et al., 1997) carried out studies to assess the relative utility of recast in Japanese and Spanish as L2. They define **corrective recast** as a response which, although communicatively oriented and focused on meaning rather than form, incidentally reformulates all or part of a learner's utterance, thus providing relevant morpho-syntactic information. Long et al. found evidence of adults' ability to learn from such implicit corrective feedback, and suggested that recast can be more effective than preemptive positive input (i.e., models of the correct form) in achieving at least short-term improvement on previously unknown L2 structures.

Aspects of communicative classroom discourse that may influence the effectiveness of recast have also been examined by Lyster (Lyster, 1998b), who studied recasts that are intended to be noticed as negative evidence (i.e., information about what is unacceptable in a given language) by young second language learners. The findings suggest that corrective feedback which is given exclusively in the form of conversational recast passes unnoticed. Students often do not recognize it as teacher correction, assuming that the teacher is responding to the content rather than to the form of their language. In another study of recast, Doughty and Varela (Doughty & Varela, 1998) investigated the effectiveness of the **corrective recasting** strategy, which is composed of two phases; repetition to draw the students' attention to the error, followed by recast to provide the target form. Their conclusions are based on empirical evidence provided by both L1 acquisition and L2 classroom studies. L1 acquisition studies, e.g. (Farrar, 1990), suggest that not only do adults provide negative evidence to children but that children notice this information and make use of it in the acquisition process. L2 classroom studies (Doughty, 1994) show that recasting behaviors arise naturally and frequently in second language classrooms, and that corrective recasting is more effective than leaving students to their own devices to develop target-like ability.

Two of Lyster's studies in L2 learning of French provide crucial evidence about the effectiveness of different types of corrective feedback. Lyster and Ranta (Lyster & Ranta, 1997) investigated how students react to different feedback strategies by studying **learner uptake** (Lyster & Ranta, 1997, p. 40), which they define as "a student's utterance that immediately follows the teachers' feedback and that constitutes a reaction in some way to the teachers' intention to draw attention to some aspect of the student's initial utterance" (Lyster & Ranta, 1997, p. 49). That is, uptake is what the student tries to do with the teacher's feedback. Two types of student uptake appeared in the data: uptake that produces an utterance still needing repair and uptake that produces a repair of the error on which the teacher's feedback focused.

Lyster and Ranta (Lyster & Ranta, 1997) examined student utterances from a corpus of 18 hours of audiotaped sessions from French as L2 classrooms. They found that approximately 34% of the student utterances contained some type of error, and that teachers responded with some type of corrective feedback to 62% of these. Of the feedback utterances produced by the teachers, 55%, were found to lead to uptake of some type on the part of the learner. However, only 27% of the feedback utterances led to student repair.

Looking more closely at the effectiveness of individual feedback strategies revealed that recast, which is the most commonly used form of feedback and is used to respond to more than half of the students' errors (55%), had the lowest rate of uptake (31%) with just over half (57%) of the reformulations being correct. Explicit correction leads to uptake only 50% of the time, but 72% of the reformulations are correct. Learner uptake for clarification requests, meta-linguistic cues, and repetition are all effective at eliciting uptake (88%, 86% and 78%, respectively), although meta-linguistic cues were found to be more successful at eliciting repair (45%) than either clarification-requests (28%) or repetition (31%). The most successful strategy for eliciting uptake was elicitation (100%), for which just under half (46%) of the students errors were repaired.

In order to get a clearer understanding of the relationships among error types, feedback types, and immediate learner repair, Lyster (Lyster, 1998a) carried out another study (also for the French language) in which elicitation, meta-linguistic cues, clarification requests, and repetition of error strategies were regrouped into a single category called **negotiation of form**, a term which more accurately captures

the ways in which teachers focused on form during meaningful interaction than the term **negotiation of meaning**, a term that had been used in prior studies (Lyster & Ranta, 1997). Negotiation of form provides learners with timely opportunities to make important form-function links in the target language without interrupting the flow of communication (Lyster, 1998a). The research findings indicate that teachers were more likely to use negotiation of form than recast in response to lexical errors, whereas grammatical and phonological errors evoked recast, but with different outcomes in terms of learner repair. Negotiation of form proved more effective at leading to immediate repair in the lexical and grammatical cases than either recast or explicit correction, whereas for phonological errors, recast was more effective.

These studies begin to shed light on the relationships among error type, feedback type, and effectiveness of feedback in L2 in French immersion classrooms. In this paper, we consider the same types of feedback strategies, looking at their frequency of use and effectiveness at eliciting repair in teaching Spanish as a foreign language. In addition, because we are interested in informing the design of ICALL systems, we analyze data from both classroom and tutorial interactions to identify the most frequent and effective feedback strategies in these two modes. Finally, because different feedback strategies may be appropriate for learners at different levels of proficiency (beginner, intermediate or advanced), we have also analyzed our data from this perspective.

Feedback in ICALL Systems for Foreign Language

ICALL systems for foreign language learning have incorporated NLP techniques to analyze learners' language production or model their knowledge of a foreign language, in order to provide learners with flexible feedback and guidance in their learning process. These systems use parsing techniques to analyze the student's response and identify errors or missing items. This allows systems to produce sophisticated types of feedback, such as meta-linguistic feedback and error reports, to correct particular student errors.

For example, as part of their ITS for German, Criswell et al. (Criswell et al., 1991) devised an advice and feedback selection component that prescribes how often and what type of feedback or advice to present. Based on error reports from an NLP module, this component determines how much information from the error report to give to the student. The system informs the student that an error was made and then either simply gives the location of the error in the sentence, or provides additional information about the nature of the error.

In subsequent work, Sams (Sams, 1995) included information about the student's errors as a type of feedback in the BRIDGE ITS, a multimedia tutoring system for German. When using BRIDGE, students receive feedback about the correctness of their responses for all exercise types. The most specific and diagnostic feedback is provided for those exercises sent to the NLP module, which analyzes the student's input to identify the specific grammatical error the student made. Students are then presented with details of their primary errors, and given a choice of viewing details about their secondary errors. This classification scheme was developed to promote language use while not constantly interrupting the intermediate-level student. The tutor could determine which errors to bring to the immediate attention of the student and which errors were less important for overall lesson objectives. In spite of this, Sams (Sams, 1995) noted that one of the main problems with the BRIDGE tutor was the way in which the error feedback was presented. He suggested that feedback should be handled as in the real world, for

example using a recast strategy. However, in order to handle feedback in this way, the system must have deeper understanding of the semantics and pragmatics of the student's utterance than was provided by the NLP component in BRIDGE.

Levin and Evans (Levin & Evans, 1995) also argue that ICALL systems for FL can benefit from NLP technology. They developed the ALICE-chan system which can identify the location of errors and explain the errors in terms of linguistic relations. This type of feedback has proven to be instructionally effective in the acquisition of case marking particles in Japanese (Nagata, 1993). However, the feedback provided by ALICE-chan is not pedagogically optimal because it uses technical terms which may be confusing to the student. Levin and Evans suggested exploring more effective approaches to the treatment of feedback in the foreign language domain.

Although ICALL systems for FL have been developed, there have been few empirical studies demonstrating the effectiveness of feedback in these systems. Nagata (Nagata, 1995) investigated the effectiveness of two types of CALL feedback: traditional feedback that indicates only missing or unexpected words in the learner's response, and feedback that provides further information about the nature of the errors in the form of meta-linguistic rules. The results of an achievement test, followed by a retention test three weeks later, showed that the second type of feedback was more effective than the first for improving the grammatical proficiency of learners of Japanese as L2 in the use of complex structures.

Nagata (Nagata, 1997) subsequently compared the effectiveness of feedback produced by an ICALL system with that of workbook instruction. The ICALL systems provided a type of elaborative and explanatory feedback based on detailed grammatical analyses of the students' attempts to produce Japanese sentences, while workbook instruction provided answer sheets for the students to check their own responses without any detailed feedback targeted to individual errors. The results of this study showed that given the same grammar notes and exercises, intelligent computer feedback was more effective than simple workbook answer sheets for developing learners' grammatical skill in producing Japanese particles and sentences.

In summary, ICALL for FL systems has shown the value of incorporating sophisticated parsers for analyzing the language produced by foreign language learners and providing them with feedback. However, the types of feedback produced by such systems take on forms not typically found in classroom interactions where natural language communication is used. In addition, it is not clear how the underlying theory of feedback and correction that implicitly guides these systems relates to the insights of SLA researchers who study what human teachers and learners actually do.

ANALYSIS OF CLASSROOM AND TUTORIAL INTERACTION

Classroom Study: Method

For the purposes of our study, we collected a corpus of classroom interactions which included 19 transcriptions of Spanish as FL classes provided by seven different teachers, totalling approximately 12 hours. Ten of the 19 classes were from Jamaica, two from Australia and six from Scotland. The teachers were distributed as follows: four from Jamaica who teach in secondary schools and whose first language is English, two Spanish native speakers from Australia who teach in a University, and one Spanish native

speaker who teaches in a further education college in Scotland. The use of 7 different teachers from three different parts of the world gives us confidence that our results are based on a balanced corpus, rather than the idiosyncrasies of a single teaching style.

The majority of students have English as their native language. The exceptions are four students from the college in Scotland, where L1 languages are French (2) and Portuguese (2). The classes transcribed include 9 at beginner level, 4 at intermediate, and 6 at advanced level. The beginner level classes were recorded in Jamaica (grades 7 and 8), the intermediate level includes classes from an Australian University (level 1 and 2) and from Jamaican Schools (grade 11). The advanced level classes are from the Scottish College.

The proficiency levels are based on those used for award of DELEs (Diplomas of Spanish as a Foreign Language), which are the official accreditation of the degree of fluency of the Spanish Language issued and recognised by the Ministry of Education, Culture and Sport of Spain. We also asked the teachers in our studies to indicate proficiency levels for each of the students in their class(es).

More details are given in Table 1.

Table 1
Classroom Transcription Data

Level	**Class**	**Teacher**	**Place**	**Focus on**	**Duration**
Beginner	Class 1	Teacher 1	Jamaica	Meaning	40’
Beginner	Class 2	Teacher 1	Jamaica	Forms	20’
Beginner	Class 7	Teacher 2	Jamaica	Forms	40’
Beginner	Class 8	Teacher 2	Jamaica	Meaning	40’
Beginner	Class 9	Teacher 3	Jamaica	Forms	20’
Beginner	Class 10	Teacher 3	Jamaica	Forms	20’
Beginner	Class 11	Teacher 3	Jamaica	Meaning	40’
Beginner	Class 16	Teacher 3	Jamaica	Forms	20’
Beginner	Class 19	Teacher 3	Jamaica	Meaning	45’
Intermediate	Class 3	Teacher 4	Jamaica	Forms	45’
Intermediate	Class 4	Teacher 4	Jamaica	Meaning	50’
Intermediate	Class 5	Teacher 5	Australia	Meaning	50’
Intermediate	Class 6	Teacher 6	Australia	Meaning	45’
Advanced	Class 12	Teacher 7	Scotland	Meaning	45’
Advanced	Class 13	Teacher 7	Scotland	Meaning	30’
Advanced	Class 14	Teacher 7	Scotland	Meaning	45’
Advanced	Class 15	Teacher 7	Scotland	Meaning	45’
Advanced	Class 17	Teacher 7	Scotland	Meaning	30’
Advanced	Class 18	Teacher 7	Scotland	Meaning	30’
Total					700’

In order to collect data, we asked teachers to record and send us samples of their natural classroom interactions. We did not instruct the teachers to focus on any particular topic or error type, or suggest the types of feedback they should use. The classroom recordings fall into two broad categories: those that focus on **forms**, that is, lessons about different grammar topics (e.g. pronominalization, syntactic

structures, possessives, passive verbs), and those that focus on **meaning**, that is, lessons in which the students discuss a variety of cultural topics in the foreign language.

Tutorial Study: Method

For the purposes of the second study, we collected a smaller corpus consisting of 5 one-on-one tutorial sessions. All tutorial sessions involved the same teacher from a Jamaican school (Teacher 3 from the classroom study), and totalled approximately 1:40 hours. All students were from the same Jamaican school and were studying at beginner level (grades 7 and 8). Both the students and the teacher have English as their native language. More details are given in Table 2.

Table 2
Tutorial Transcription Data

Level	**Class**	**Teacher**	**Place**	**Focus on**	**Duration**
Beginner	Class 2T	Teacher 3	Jamaica	Forms	20'
Beginner	Class 3T	Teacher 3	Jamaica	Meaning	20'
Beginner	Class 4T	Teacher 3	Jamaica	Meaning	20'
Beginner	Class 5T	Teacher 3	Jamaica	Forms	25'
Beginner	Class 6T	Teacher 3	Jamaica	Meaning	15'
Total					100'

Data Analysis

Our aim is to determine how teachers give effective feedback and/or acknowledgement to the student. We are interested in when, that is, under what conditions, teachers explicitly acknowledge the student's contribution and when they simply continue with the next question, topic or explanation. We are also interested in how teachers acknowledge the student in different circumstances, and what type of corrective feedback is most frequently and effectively used. Better understanding of these factors will help us to inform the design of ICALL systems for FL.

To this end, we devised a scheme for annotating the dialogue moves of both students and teachers in FL teaching interactions. Our scheme draws on work by Core et al. (Core et al., 2002), who developed an annotation scheme for tutorial dialogue to identify the communicative acts that make up dialogue strategies, as well as patterns of interaction, and their frequency. Core et al.'s scheme is based on DAMSL (Allen & Core, 1997), which was initially developed for annotating task-oriented dialogues. Following DAMSL, Core et al. have adopted the notion that utterances in a dialogue can have both forward-looking and backward-looking functions in the dialogue. Forward-looking functions indicated how the current utterance constrains the future beliefs and actions of the participants, and affects the discourse. These include questions, such as Diagnostic-query (testing student knowledge) and Information-request (getting information), requests, such as Action-directive (suggestions/request for actions), and

Statements. Backward-looking functions indicate how the current utterance relates to the previous discourse. These include Accept (accepting a statement, request or answer), Reject (rejecting a statement, request, or answer), Follow-up (a question asking for more details).

Core et al.'s scheme also considers question structure, which identifies the point at which a question is answered as well as who answered it, and correctness. Correctness labels indicate whether the student's answer is correct, partially correct, and so forth. We needed to add only three correctness labels to cover typical student answers in the FL teaching context. However, since these detailed distinctions are not used in the analyses we report in this paper, we will not discuss them here for purposes of brevity. For a full description of the annotations we have performed on our corpora, see (Ferreira-Cabrera, 2003).

Core et al. do not define categories for the types of teacher feedback moves that are characteristic of FL teaching, and thus we had to develop our own feedback classification scheme for teacher utterances. We discuss these next.

Feedback and Error Tags

In order to determine the types of feedback that are most frequently used by the teachers in classrooms, and the contexts in which they are most effective, we annotated our dialogues with types of feedback and types of error.

Following the distinctions made in the section on positive and corrective feedback strategies, positive feedback was annotated as **repetition**, **rephrasing**, **acknowledgement** or **accept**. Corrective feedback was annotated with either a GAS label (**repetition**, **correction**, **recast** or **give answer**) or a PAS label (**meta-linguistic cue**, **clarification** or **elicitation**).

For the purpose of our study and considering previous pilot analyses, we also established specific types of grammar, vocabulary and pronunciation errors:

- **Grammar** errors: we found the following most frequent types of grammatical errors:

    1. **Structure:** Errors in the syntactic structure or word order.
    2. **Closed classes:** Errors in the use of a pronoun (pro), determiner (det), preposition (prep), etc.
    3. **Conjugation:** Errors in the conjugation of verbs.
    4. **Agreement:** Errors in agreement between subject/verb, determiner/noun, noun/adjective.
    5. **Tense:** Errors in verb tense.

- **Vocabulary** errors[3]:

    1. Unaware of the term.
    2. Inappropriate or inaccurate use of Spanish term.

[3]This corresponds to Lyster's **lexical** error category.

- **Pronunciation** errors:

    1. **Accent:** Inappropriate accentuation of Spanish word(s).
    2. **Mispronunciation:** of Spanish word(s).

Transcriptions and Data Processing in XML

The classroom recordings were transcribed by a native Spanish speaker, and annotated with the labels defined in the tutorial annotation scheme described above. We used an annotation tool that inserts XML tags into the data files, so that they can then be processed using a variety of XML tools, including tools for querying and viewing the data. For this study, we focused on the labels that will shed most light on our research questions. In particular, we were interested in the following:

- Type and frequency of grammar, vocabulary and pronunciation errors.
- Type and frequency of different types of feedback (GAS, PAS, repetition, rephrasing).
- Type and frequency of feedback after grammar, pronunciation and vocabulary errors.
- Type of student utterance 1, 2 and 3 turns after PAS, GAS and correct answer.

RESULTS

Frequencies of types of positive feedback

To answer our first research question, we investigated the type of feedback that teachers give to the students when they answer correctly. The results in Table 3 show broadly the same pattern for both teaching modes. Repetition or rephrasing is used about 40% of the time and simple acknowledgement or acceptance about 30% of the time. For the remaining 30% of cases, there was no feedback. In these cases there was a change of topic or another question. Of the two feedback strategies highlighted in the section on positive and corrective feedback strategies, repetition was most frequently used when the teacher chose to do more than simply accept or acknowledge.

Table 3
Positive-Feedback distributed by Teaching Mode

Feedback	Tutorial	Classroom
Repetition	49 (33%)	127 (27%)
Rephrasing	12 (8%)	49 (10%)
Acknowledge	10 (7%)	18 (4%)
Accept	34 (23%)	123 (26%)
Total	105 (71%)	317 (67%)
# Correct	147	471

Frequencies of types of corrective feedback

To shed light on our second research question, we considered the distribution of the different types of errors and corrective-feedback in the two teaching modes.

Table 4
Classroom Error Frequency by Learning Level (**Gram**mar, **Voc**abulary, **Pro**nunciation)

Level	**Gram**	**Voc**	**Pro**	Total
Beginner	66 (47%)	9 (6%)	66 (47%)	141
Intermediate	49 (59%)	28 (34%)	6 (7%)	83
Advanced	28 (45%)	23 (37%)	11 (18%)	62
All levels	143 (50%)	60 (21%)	83 (29%)	286

First, we focus on the distribution of errors. Some students' utterances contain more than one error. However, we found that teachers typically treated just one error at a time in order to keep a fluent conversation going.

Table 4 shows the errors that were made in the classroom study, broken down by learning level. Here we see that the relative frequencies of the different types of error differ according to the level of the learner. Grammatical errors are among the most frequent for learners at all levels. For beginners, pronunciation errors are equally as frequent as grammatical errors, while vocabulary errors are the next most prevalent for intermediate and advanced learners.

Table 5
Error Frequency by Teaching Mode

Error	**Tutorial**	**Classroom**
Grammar	48 (35%)	143 (50%)
Vocabulary	20 (15%)	60 (21%)
Pronunciation	69 (50%)	83 (29%)
Total	137	286

Table 5 shows the errors by type and teaching mode. This table demonstrates two interesting differences in the total errors of the two teaching modes. First, although it appears that a smaller number of errors occurred in tutorial mode than in classroom mode, when we take into account that we have 700 minutes of classroom data and only 100 minutes of tutorial data, we see that in fact errors were significantly less frequent in the classrooms. This may be due to the fact that in tutorial mode the student has more opportunity to interact with the teacher than in classroom mode.

The second difference relates to the frequency of different types of errors in the two modes. From Table 5, we see that in classroom mode grammar errors were by far the most frequently occurring type, whereas in tutorial mode, pronunciation errors were the most common. This difference may be attributable to the fact that in tutorial mode students have more opportunities to speak, and therefore have more opportunities to make pronunciation errors, than students in classroom mode. This finding may also

be explained by the fact that students are more likely to volunteer in the classroom when they know the answer, while in tutorial mode the individual student has to answer every question posed to them. With more data, it might be worth investigating further the difference in errors in tutorial mode and beginning level classroom mode.

Table 6
Corrective feedback Frequency by Teaching Mode

Feedback	**Tutorial**	**Classroom**
GAS	101 (74%)	177 (62%)
PAS	27 (20%)	68 (24%)
No feedback	9 (6%)	41 (14%)
Total Errors	137	286

Now we turn to the frequency of different types of corrective feedback in the two modes. Table 6, shows the number of errors that received GAS, PAS or no feedback in the two instructional modes. This shows that in the interactions we studied, the majority of student errors received some kind of feedback. Indeed, our figures for both tutorials and classrooms exceed the 62% found by Lyster and Ranta (1997). From the table, we also see that GAS are much more frequent than PAS in both modes. That is, regardless of the mode of interaction, teachers preferred to handle the students' errors with corrective feedback in which the teacher gives the location of the error or provides the target form. However, if we aggregate the PAS and GAS feedback in each mode, we see that there is a significant difference in error correction in the two modes ($\chi^2(2) = 5.36, p < 0.025$). In particular, errors received no feedback in classroom mode significantly more frequently than in tutorial mode. In cases where errors did not receive feedback, the errors were followed by a move whose DAMSL annotation was Reject, Follow-up, Diagnostic-query, or Change of topic. Prior studies of the treatment of learners' errors in classrooms have shown that teachers do not treat all of the errors that occur during the classroom session (Allwright, 1975; Chaudron, 1977; Fanselow, 1977; Long, 1977). The fact that errors are more likely to receive corrective feedback in tutorial mode may be due to several factors. In classrooms there is time pressure and the need to keep the conversation moving. Perhaps the teacher needs to weigh whether or not the whole class would recognize the error to judge whether the non-correction would cause difficulties. Also, the teacher must weigh up whether the correction will benefit most of the class as opposed to just the individual. Furthermore, the difference may also be due to the nature of one-on-one dialogue, where utterances that are not signalled as problematic are taken as accepted (Clark & Schaefer, 1989).

Table 7 shows the frequencies of GAS and PAS for each type of error. When we consider the total number of uses of GAS and PAS feedback, i.e., combining the counts from classroom and tutorial mode, there is a significant interaction between feedback type and error type ($\chi^2(2) = 76.79, p < 0.0001$), suggesting that the type of error affected the choice of GAS versus PAS. However, this is due to the fact that regardless of mode, virtually all pronunciation errors received some type of GAS. If we consider only grammar and vocabulary errors, there is not a significant correlation between error type and feedback type. Teachers use GAS frequently to correct grammar and vocabulary errors in both modes. In fact,

Table 7
Rate of GAS and PAS per Error type (**Gram**mar, **Voc**abulary, **Pron**unciation)

Error	GAS		
	Tutorial	**Classroom**	**Total**
Gram	24/48 (50%)	74/143 (52%)	98
Voc	13/20 (65%)	23/60 (38%)	36
Pro	64/69 (93%)	80/83 (96%)	144
Total	101/137 (74%)	177/286 (62%)	278

Error	PAS		
	Tutorial	**Classroom**	**Total**
Gram	21/48 (44%)	49/143 (34%)	70
Voc	5/20 (25%)	19/60 (32%)	24
Pro	1/69 (1%)	0/83 (0%)	1
Total	27/137 (18%)	68/286 (24%)	95

GAS were used for a higher percentage of errors than PAS in all cases.

Effectiveness of corrective feedback strategies

Thus far, we have observed that for corrective feedback, GAS are more frequently used than PAS for all three types of errors (grammar, vocabulary and pronunciation). But frequency alone does not imply that these strategies are the most effective. Thus, for each type of error, and in each mode of interest, we must determine whether GAS is correlated with student learning, and whether GAS lead to higher learning gains than PAS (our third research question).

To assess learning gain, we analysed what happens in the three student turns immediately following the teacher's corrective feedback to the student's incorrect answer. We considered whether or not the student repaired the error. By "repair", we mean the correct reformulation of an error as uttered in a single student turn after a GAS or PAS. In cases where the student's response to the feedback was a correct answer, we counted this as one repair, but in cases where the utterance of the student still needed repair, we did not count. If the teacher used more that one feedback strategy for correcting one error, we just associated the repair with the strategy that prompted the repair. For example in the following sequence, we just counted a repair for the GAS strategy:

S: Makes an error

T: Gives a PAS Strategy

S: Repeats error or no response

T: Gives a GAS Strategy

S: Responds with a correct answer

We believe that the relationship between the frequency of repairs and frequency of each type of corrective feedback gives an indication of the immediate effectiveness of the feedback, at least in terms of what the student did or did not try to do with the feedback. We think this may indicate that the student has noticed the error and the correct answer may indicate "a step at least toward acquisition". According to Lightbown (Lightbown, 1998) "The fact that the learner does not make an immediate behavioral change cannot be taken as evidence that there is no effect of (some corrective feedback operationalized as) 'focus on form'. Nor can a corrected response from the learner be taken as evidence that a more correct or advanced form has been integrated into the learner's inter-language. Nevertheless, a reformulated utterance from the learner gives some reason to believe that the mismatch between learner utterance and target utterance has been noticed, a step at least toward acquisition". This awareness of the gap between what students want to say and what they can say, and what they do know and what they do not know, as only partially provided by some corrective strategies, can be a first step towards improvement.

Table 8
Rate of GAS and PAS Repair in Classroom and Tutorial Mode

Classroom Error Type	GAS	G-Repair	PAS	P-Repair
Grammar	74	36/74 (48%)	49	39/49 (80%)
Vocabulary	23	9/23 (39%)	19	18/19 (95%)
Pronunciation	80	67/80 (83%)	0	0
Total	177	112/177 (63%)	68	57/68 (84%)

Tutorial Error Type	GAS	G-Repair	PAS	P-Repair
Grammar	24	13/24 (54%)	21	16/21 (76%)
Vocabulary	13	3/13 (23%)	5	4/5 (80%)
Pronunciation	64	56/64 (87%)	1	0/1 (0%)
Total	101	72/101 (72%)	27	20/27 (74%)

The top part of Table 8 shows the rate of GAS and PAS repair in classroom mode. Overall, the proportion of repairs after PAS was higher than that after GAS. A statistical analysis considering the number of repaired and non-repaired errors following GAS and PAS for each type of error showed that for grammar ($\chi^2 = 11.86, p < 0.0005$) and vocabulary errors ($\chi^2 = 14.01, p < 0.0001$), PAS are more effective than GAS. For pronunciation errors, GAS are the only strategies used and they elicit a high rate of repair.

The lower part of Table 8 shows that there is a similar pattern in tutorial mode. When we look at effectiveness broken down by error type, we see that a higher percentage of grammar errors are repaired after PAS than after GAS, but the results are not significant. For vocabulary errors, PAS are more effective than GAS ($\chi^2 = 4.92, p < 0.05$). Finally, virtually all pronunciation errors invite GAS feedback, and there is a high rate of repair.

In summary, these results indicate that PAS are more effective than GAS in eliciting repair for the errors they were used to treat. Despite this, teachers use GAS more frequently for all error types in both classroom and tutorial mode.

Effectiveness of Individual Strategies by Learner Level

For the purposes of informing the design of ICALL systems for FL, it is important to identify the factors that indicate which strategy to use for corrective feedback in a given situation. We have already seen that the type of error is one factor that influences this decision. Lyster and Ranta (Lyster & Ranta, 1997) argue that the learner's level of proficiency in the target language is another important factor that teachers should take into account when choosing a feedback strategy. Thus, we analyzed our data broken down by learner level as well as the type of error.

Table 9
Proportion of Repair of GAS and PAS by Learner Level

Level	**Beginner**	**Intermediate**	**Advanced**	Total
G-Repair	82/110 (75%)	15/36 (42%)	15/31 (48%)	112
P-Repair	25/27 (93%)	22/32 (69%)	9/9 (100%)	56
All Strategies	107/137 (78%)	37/68 (54%)	24/40 (60%)	168

As we can see from Table 9, PAS are more effective than GAS for learners at all levels. There is a problem with breaking this down further by error type and feedback strategy, because the amount of data does not always allow statistically significant conclusions to be drawn. However, the following are the main trends that we observed.

First for the treatment of grammar errors, among GAS, recast (the most frequent strategy found by (Lyster & Ranta, 1997)) was the most frequent strategy at beginner and intermediate levels, whereas explicit correction was the most frequent strategy at the advanced level. However, recast led to repair just 36% of the time overall. Give answer evoked a high rate of repair among learners at beginner and intermediate levels, but was not used to correct grammar errors of advanced learners. Explicit correction was used at all levels, and evoked a high rate of repair from intermediate students, but was less successful with beginner and advanced learners. Finally, there are too few cases of repetition to draw any conclusions.

Among the PAS strategies for grammar errors, elicitation is the most frequently used strategy, however, whereas (Lyster & Ranta, 1997) found this to be the most effective strategy, we observed it leading to repairs less often than metalinguistic cues or clarification requests. Metalinguistic cues are extremely effective for beginning and advanced students, but less so for learners at the intermediate level. A more detailed look at the instances of meta-linguistic prompts showed that they were especially effective for improving aspects of grammar (i.e., the subjunctive mood) in high-intermediate and advanced levels. Our corpus did not include any instances of clarification requests for beginners, but this strategy always led to repair in the few cases where it was used for intermediate and advanced learners.

For vocabulary errors, the situation is less clear due to the small number of occurrences of this type of error in our corpora. Overall, although recast was the most frequently used strategy of the GAS group, it yielded only 27% repaired errors. Explicit correction was the most effective strategy with 71% of the errors repaired. The results for the PAS group indicate that all strategies are used with similar frequency,

and that all strategies evoke high rates of repair. This is consistent with the high rates of uptake for PAS strategies found by (Lyster & Ranta, 1997).

For pronunciation errors, we focus on the strategies in the GAS group because the overwhelming majority of these errors were corrected by GAS. Overall, explicit correction is the most frequently used strategy and leads to repair 85% of the time. Give answer and recast were used with similar frequency, but give answer was much more effective at eliciting repairs (92% vs. 58%). At the beginner and advanced levels, give answer and explicit correction were the most effective GAS strategies for treating pronunciation errors. There were too few repairs at the intermediate level to draw any conclusions.

The results of this and the previous section provide useful guidance for the design of feedback in ICALL systems, which we discuss further in the section on implications for the design of ICALL systems.

INEFFECTIVE GAS AND PAS

The relationship between the frequency of repairs and frequency of each type of corrective feedback gives an indication of the immediate effectiveness of the feedback. In cases where the student's response to the feedback is a correct answer, this may indicate that the student has noticed the error and has made a step toward acquisition. However, there are cases in our corpora in which the student either did not notice the target form provided by the teacher's feedback or did not have the knowledge that the teacher was pushing for.

The effectiveness of corrective feedback depends on the particular aspects of language being corrected as well as the conditions relating to the provision of the feedback. In this study, we investigated two such conditions. The first is the opportunity, or lack thereof, for the student to respond to the corrective feedback. We found that in a non-negligible percentage of cases, after providing feedback, the teacher went on to the next question, topic or another type of feedback (clarification, explanation) without waiting for the student to produce an answer or confirm the feedback. This is the concept known as **wait time** (Rowe, 1969), that is, the length of time that a teacher waits after asking a question of a learner, before prompting, rephrasing or redirecting the question.[4] We believe this concept is important not only for response to questions but also for response to corrective feedback. If the teacher were to wait longer after providing corrective feedback, the student might reply and provide a repair.

The second condition is the learner's ability to produce the correct answer. In the case of GAS feedback, where the target form is typically provided, students may not notice a gap between their use of non-target forms and the teacher's corrected form (Schmidt & Frota, 1986; Swain, 1985). Corrective feedback, and in particular recast, does not always trigger students to notice that they have made an error. Lyster and Ranta (Lyster & Ranta, 1997) point out that recasts are called "echoes" in French, even though they do not exactly repeat what the student has said, because the learners often do not perceive

[4]Rowe (1969) studied "wait-time" with native English-speaking children studying science. She found that as teachers increased their wait-time, the quality and quantity of the students' answers increased. In the same way, Holley and King (1974) asked teachers of German as L2 to wait five to ten seconds if a learner erred or hesitated in answering a question. They reported that in over 50% of the cases no corrective efforts from the teacher were needed. The students themselves were able to respond correctly, given this brief additional pause (Rowe, 1969; Holley & King, 1974; Fanselow, 1977; Allwright & Bailey, 1991).

the difference between their incorrect form and the teacher's reformulation. Students may not recognize the teacher's correction because they assume that the teacher is responding to the content rather than to the form of their language. Indeed, as we saw in Table 3, when giving positive feedback, teachers often repeat or rephrase students' well-formed utterances to reinforce or build on what students' have said.

We now discuss these two issues in more detail.

Lack of opportunity for student response after feedback

In order to get a clearer idea about what happened with the ineffective GAS and PAS feedback, we investigated all GAS and PAS turns which did not lead to repairs in our corpus. One explanation of the failures can be drawn from Chi et al. (Chi et al., 2001), who propose an interaction hypothesis (I-hypothesis). The I-hypothesis essentially states that the effectiveness of tutoring arises from the close interaction of the student and tutor. We define "interactive" (**Int**) feedback to be those instances of feedback that obtained a response from the student, and "Non-interactive" (**Non-Int**) feedback as cases in which the tutor gives feedback followed immediately by another question, a topic change, or another feedback move (explanation, clarification) without giving the student a chance to respond.

Table 10
Interaction in Tutorial and Classroom Modes (one teacher)

Tutorial	**GAS**	**Int**	**Non-Int**	**PAS**	**Int**	**Non-int**
Gram	24	16 (67%)	8 (33%)	21	21 (100%)	0
Voc	13	3 (23%)	10 (77%)	5	5 (100%)	0
Pro	64	62 (97%)	2 (3%)	1	1 (100%)	0
Total	101	81 (80%)	20 (20%)	27	27 (100%)	0

Classroom	**GAS**	**Int**	**Non-Int**	**PAS**	**Int**	**Non-Int**
Gram	18	17 (95%)	1 (5%)	9	9 (100%)	0
Voc	2	2 (100%)	0	2	2 (100%)	0
Pro	24	23 (96%)	1 (4%)	0	0	0
Total	44	42 (95%)	2 (5%)	11	11 (100%)	0

Table 10 shows the percentages and total number of feedback moves in which the teacher did not give students the opportunity to repair their error (**Non-Int**). Here we only include the classroom data from the same teacher as in the tutorial interactions. In tutorial mode, we see that of a total of 27 PAS all were interactive, that is, all PAS were followed by a student response. In contrast, of a total of 101 GAS, 20 were followed by a non-interactive turn. That is, in 20% of the cases where GAS were used, the student was not given the opportunity to produce a response. In some cases this occurs because the teacher appears to give priority to the flow of conversation and does not want to disrupt the communication.

Now consider the results for classroom mode shown in the lower part of Table 10, keeping in mind that this is a reduced dataset taken only from the same teacher who participated in the tutorial interactions. Here we see that in the classroom interactions, there were only two cases of GAS (5%)

in which the teacher did not give the student the opportunity to answer. This implies that for the GAS group, non-interactive feedback is less of a problem in classroom mode than in tutorial mode, at least for this particular teacher. As for PAS, just as in tutorial mode, there are no cases in which PAS is non-interactive.

In brief, our analysis concerning interactive vs. non-interactive corrective feedback showed one interesting trend: The PAS group is more interactive than the GAS group in both teaching modes. This confirms our assumption that the strategies we have labelled PAS really are used for "prompting the answer" in practice. It may also help to explain why students repair less with GAS than PAS - because somehow PAS "demand" a reply, whereas GAS may signal that a repair is not needed.

Student Inability to Notice the Target Form or Construct the Correct Answer after Feedback

We also investigated what happened with the rest of the ineffective GAS and PAS strategies by examining all turns of GAS and PAS in our corpus that did not elicit repairs, even though they were interactive. Chi and colleagues (2001) proposed that the interactive nature of students' responses needs to be differentiated from students' constructive responses in order to tease apart the independent contributions of being constructive from being interactive. They differentiated between responses that make a substantive contribution with respect to the content versus those that do not.

In general, we note that in both modes, there were a number of cases in which the student did not answer correctly after feedback, that is, their answer did not constitute a substantive contribution with respect to the given feedback. This may be due either to the fact that they did not have the knowledge necessary to construct their own correct answer, or they did not perceive the teacher's correction as such. We considered both of these as an interactive but non-constructive type of exchange. That is, a student can be interactive in the sense of giving an answer and yet be non-constructive in terms of the correctness of the answer.

Table 11 shows that in tutorial mode, of a total of 81 interactive GAS (Int-GAS), 9 (11%) were not repaired (No G-Rep) by the student. With respect to the PAS group, it shows that 26% of interactive PAS moves in tutorial mode were not repaired by the student. In this case, the problem seems to arise because the student does not know the answer to the question. For classroom mode (considering the one teacher only), the second table shows that in 8 out of 42 instances following GAS (19%), students were unable to repair their errors.

To determine what we have learned from our analysis of interaction and construction, we now compare the data for overall rates of repair for PAS and GAS (*what percentage of PAS/GAS led to repair?*), as presented in Table 8, with our data for interactive GAS and PAS repair rates (*what percentage of interactive PAS/GAS led to repair?*) We combined this data in Table 12 for the convenience of the reader. Here we show repair rates for all error types, as well as repair rates for grammar and vocabulary errors only. What we see in both cases is that interactive repair rates are considerably higher than overall repair rates. Moreover, this difference is much greater for GAS feedback than PAS feedback in both modes. In addition, the differences are more striking when we remove pronunciation errors from the counts. For grammar and vocabulary errors, interactive GAS repair exceeds overall GAS repair by 28%

Table 11
Unrepaired Errors in Tutorial and Classroom Modes

Tut. Error	Int-GAS	G-Rep	No G-Rep	Int-PAS	P-Rep	No P-Rep
Gram	16	13	3 (19%)	21	16	5 (24%)
Voc	3	3	0	5	4	1 (20%)
Pro	62	56	6 (10%)	1	0	1 (100%)
Total	81	72	9 (11%)	27	20	7 (26%)

Class. Error	Int-GAS	G-Rep	No G-Rep	Int-PAS	P-Rep	No P-Rep
Gram	17	12	5 (29%)	9	8	1
Voc	2	2	0	2	2	0
Pro	23	20	3 (13%)	0	0	0
Total	42	34	8 (19%)	11	10	1 (8%)

in classrooms and 41% in tutorial mode. Again, the differences for PAS are much smaller.

Of course, it is not surprising that interactive repair rates are higher than overall repair rates, because of our definition of interaction. However, the magnitude of the differences indicates that there is a great benefit to be gained from interaction, especially in the case of GAS feedback for grammar and vocabulary errors. We discuss this and its relation to wait time further in our concluding remarks.

DISCUSSION

The two studies we have done have helped us to shed light on our research questions and have furthered our ultimate aim of providing guidance for the design of feedback strategies for ICALL systems for Spanish as a foreign language. In terms of our research questions, we have noted the following:

1. *What are the most frequent types of positive feedback used by teachers of Spanish as FL after a student's correct answer?* For both teaching modes, we found a similar pattern. Repetition or rephrasing is used roughly 40% of the time, and simple acknowledgement or acceptance about 30% of the time. In the remaining 30% of cases, no explicit positive feedback was given, and the teacher moved on with a change of topic or another question. Repetition is used roughly three times more frequently than rephrasing in both modes, whereas accept is used three times more frequently than acknowledgement in tutorial mode, but 6 times more frequently in tutorial mode (Table 3), but this difference is not significant.
2. *What are the most frequent types of corrective feedback used by teachers in Spanish as FL, and how do they relate to learner errors? In particular, what are the relative frequencies of GAS and PAS?* We have seen that teachers use GAS more frequently for all types of errors (pronunciation,

Table 12
Overall vs. Interactive Repair in Classroom and Tutorial Mode

Interactive Repair: All Error Types		
	Int-GAS-Repair	**Int-PAS-Repair**
Classroom	81%	91%
Tutorial	89%	74%
Interactive Repair: Grammar and Vocabulary Errors		
	Int-GAS-Repair	**Int-PAS-Repair**
Classroom	74%	91%
Tutorial	84%	77%
Overall Repair: All Error Types		
	All-GAS-Repair	**All-PAS-Repair**
Classroom	63%	84%
Tutorial	72%	74%
Overall Repair: Grammar and Vocabulary Errors		
	All-GAS-Repair	**All-PAS-Repair**
Classroom	46%	84%
Tutorial	43%	77%

grammar and vocabulary), even though these strategies elicit a low rate of repair for grammar and vocabulary errors (Table 8).

3. *What are the most effective types of corrective feedback? In particular, are there differences in the apparent effectiveness of GAS and PAS?* Despite the fact that teachers use GAS more frequently than PAS, examination of the relationship between corrective feedback and student repair indicates that PAS are more effective than GAS for the treatment of grammar and vocabulary errors (Table 8). However, vocabulary errors occurred with low frequency in our corpora, and thus further empirical investigation is required. The situation is different for pronunciation errors, which almost exclusively elicit GAS feedback from the teacher, and GAS frequently leads to repair (Table 8). However, note that the repair is typically not self-repair, but is of the type that Lyster and Ranta (Lyster & Ranta, 1997) call *repetition*, because the teacher provides the target form and the student repeats it.
4. *Are there differences in type, frequency, and/or effectiveness of the strategies in the two teaching modes?* The frequency of use and rates of repair for GAS and PAS feedback follow the same pattern in the two modes. GAS are used considerably more frequently than PAS (Table 6), and used almost exclusively for pronunciation errors in both modes. Although the frequencies of repair differ in the two modes, they follow a similar trend: PAS are more effective in terms of eliciting repair of grammar and vocabulary errors, and GAS elicit a high rate of repair for pronunciation errors (Table 8).

   We did find differences between the two modes in terms of interaction and constructive repair. Although PAS, by their nature, are virtually always interactive in both modes, GAS are more interactive in classroom mode than tutorial mode (Table 10). In addition, students were unable to

respond constructively (correctly) after GAS feedback more frequently in classroom mode (19%) than in tutorial mode (11%). Interestingly, this trend is reversed for PAS feedback, where students failed to correct their errors in 26% of the cases in tutorial mode, but in only one case in classroom mode. Although there were only 11 cases of interactive PAS feedback in classroom mode, and thus we must study larger corpora before drawing conclusions, it is interesting to speculate as to why this may be the case. One explanation has to do with affect. Perhaps teachers feel that they can push students more in one-on-one tutorial interactions, where students will not be embarrassed in front of their peers if they do not know the correct answer.

Compared to (Lyster & Ranta, 1997), our corpora showed a higher rate of teachers using corrective feedback and generally higher effectiveness (though different measures were used). Our results confirm their conclusions about the high effectiveness of PAS strategies and the fact that the most frequent strategies used (in their case, recast; in ours, GAS in general) are not always the most effective. Our results also shed light on differences arising from error type and teaching mode.

There are a number of reasons why teachers may prefer to use GAS feedback. First, in classrooms where the focus is on communication, teachers seem to be concerned about maintaining the flow of the conversation. They try to facilitate students' error resolution by providing them with the target forms, rather than encouraging self-repair sequences which may disrupt the flow of the conversation. Second, students who are not yet proficient in the target language may not have the knowledge needed to self-repair, and therefore teachers must rely on modelling techniques such as recast and explicit correction. Third, teachers may use GAS feedback because they are concerned about letting errors go uncorrected, which could mislead not only the student making the error, but also the other students (at least in classroom mode) and cause them to form misconceptions.

Although GAS feedback is by far the most frequently used for all error types, we have seen that it is less effective for grammar and vocabulary errors than the strategies in the PAS group. One reason for this may be that students simply do not perceive GAS feedback as teacher correction, that is they do not notice the difference between their erroneous utterance and the teacher's reformulation. For example, Mackey et al. (Mackey et al., 1990) found that the use of recast for morpho-syntactic errors led to a low rate of repair, and they argue that this is because the feedback was generally not perceived as correcting morpho-syntactic errors, and therefore it provided little opportunity for uptake. In classrooms that are focused on a communicative approach to FL teaching with an emphasis on meaning, such as many of those we studied, students often assume that the teacher is responding to the content rather than to the form of their language. This is not surprising given the findings of Lyster and Ranta (Lyster & Ranta, 1997), who reported that teachers in their study frequently repeated students' well-formed utterances to reinforce what students have said or to build further on students' contributions. This ambiguous use of recast could only serve to exacerbate the perception problem from the learner's perspective.

With regard to repair, there seems to be agreement in the pedagogical literature that self-repair (i.e., students correcting their own linguistic errors) is more effective for learning than teacher-repair for several reasons. First, self-repair leads to knowledge construction as students actively confront errors and extend or revise their hypotheses about the target language (Swain, 1995; Lyster & Ranta, 1997). Second, self-repair allows students to automatize retrieval of the target language knowledge they already

have (Lyster & Ranta, 1997). And finally, self-repair may be more conducive to acquisition than other-repair because it is less likely to result in a negative affective response (Van Lier, 1988). Therefore, when considering the effectiveness of feedback strategies, it is important to distinguish other-initiated self-repair (teacher notices student's error and prompts repair but the student generates the correct form him/herself) from other-initiated other-repair (teacher notices the student's error and the teacher corrects the error).

Feedback strategies in our GAS group tended to produce other-initiated other-repairs in which the teacher corrects the error. This is not surprising since three of the four GAS strategies (explicit correction, recast, and give answer) by definition provide the student with the target form and student repairs come in the form of repetition. On the other hand, feedback strategies in our PAS group elicit other-initiated self-repair, where the teacher prompts the student for a correct answer and the student must try to provide it. Therefore our results showing the superior effectiveness of the strategies in our PAS group provide empirical evidence (in the domain of learning Spanish as a foreign language) for Chi et al.'s (Chi et al., 2001) claim that substantive construction from interaction is important for learning. In addition, our results concur with those of Lyster and Ranta (Lyster & Ranta, 1997), and this indicates that their results are not particular to French immersion instruction, but are more generally applicable.

IMPLICATIONS FOR THE DESIGN OF ICALL SYSTEMS

Our studies indicate that PAS strategies effectively support the learning process for grammar and vocabulary, suggesting that ICALL systems should implement ways to prompt students to construct responses.

As discussed earlier, ICALL systems have incorporated parsing technology to analyze the students' responses and identify errors or missing items, thus allowing these systems to produce sophisticated types of feedback, such as meta-linguistic explanations and "error reports", to correct particular student errors. However, as we noted, these types of feedback often take the form of unnatural corrections, which tend to disrupt the natural flow of conversation that typifies the current teaching approaches that SLA researchers advocate (Criswell et al., 1991; Sams, 1995; Levin & Evans, 1995; Nagata, 1995; Nagata, 1997). Although meta-linguistic feedback is a suitable strategy for intermediate and advanced students (Sorace, 1985), and is a very frequently used strategy in the PAS group, we observed that in the majority of the cases in our data, this feedback takes a simpler structure than the feedback strategies employed in previous ICALL systems (Nagata, 1995; Nagata, 1997). Moreover, in human-human interactions, the feedback strategies in our PAS group are more effective at eliciting repair than the more frequently used strategies from the GAS group, and thus we believe it is crucial to design ICALL systems for foreign language learning that implement our results, in order to evaluate their effectiveness and to compare them to previous approaches to feedback in ICALL systems.

The implementation of strategies in our PAS group requires that the ICALL system be able to carry on an appropriate interaction with the student. Although unconstrained conversation of the type that human teachers employ is beyond reach, recent advances in tutorial dialogue systems research make the interactive techniques we propose more feasible than they were at the time many ICALL systems were designed. This research shows that sophisticated interactions can be carried on in domains for which rich underlying models have been developed (Bratt et al., 2005; Zinn et al., 2005), or for which possible cor-

rect and incorrect responses have been enumerated and feedback moves for each case have been authored (Graesser et al., 2004; Rosé et al., 2001). In addition, recent work has shown that an ITS in which students produced self explanations by making selections from a menu led to learning outcomes that were equivalent to a version of the system in which students explained their problem-solving steps in their own words (Aleven et al., 2003). This result suggests that full-blown natural language understanding may not be required in order to support interactions that evoke knowledge construction.

A Decision-making Model for Feedback Moves in ITS

Our results indicate that in order to determine what type of feedback to give, a system must take into account: the type of error the learner has made (grammar, vocabulary or pronunciation error), and the learner's level of proficiency (beginner, intermediate, advanced). Although in many cases we do not have enough data to reach statistical significance when we break down the repairs by error type and individual feedback strategies, the trends we have noted give us a starting point for performing more controlled experiments and hints at how we might formulate an initial algorithm for feedback generation. The next step is to build a parameterized system that can be systematically manipulated to do a range of controlled studies to determine which feedback strategies are most effective for each error type.

We have defined a model for the design of a feedback component for ICALL systems for Spanish as a foreign language (Figure 1), which takes the two factors into account (the type of error and the learner's level of proficiency). In our model, we assume that error analysis is performed by an interpreter/analyzer. As noted above, prior ICALL systems have made successful use of parsing technology to identify grammar errors, and recent research on a reading tutor has shown that given good expectations about what the student is trying to say, automatic speech recognition can be used to identify pronunciation errors (Tam et al., 2003). In this model, we also assume that the learner's level of proficiency is given.

In our model, the feedback sequence starts when a student's answer contains at least one error. If the answer contains more than one error, the system must determine which error should be treated first, and in our model this decision is based on the learner level. For beginners, grammar and pronunciation errors are the most frequent, and thus we suggest that priority should be given to the treatment of these types of errors. For intermediate and advanced learners, grammar and vocabulary errors should be addressed first.

Once an error is identified, a feedback strategy must be chosen, as shown in the model for feedback generation in Figure 2. To determine whether to use GAS or PAS feedback after the first error, we propose the following general algorithm. In this, the choice of GAS vs. PAS is motivated by the results in Table 8 and the choice of individual strategies is motivated by the finer analysis reported in the section on effectiveness of individual strategies by learner level, as well as the desire to provide interesting variation in the strategies used.

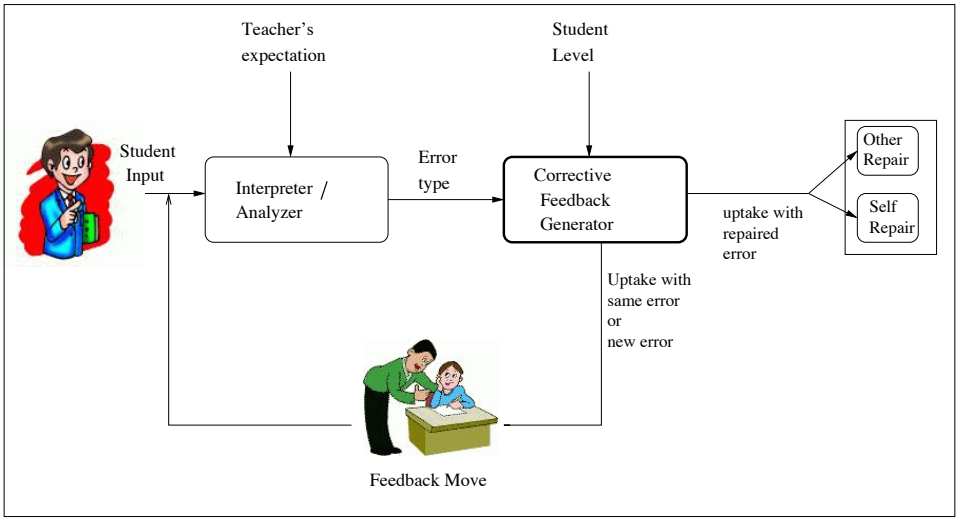

Fig.1. The Process of Error Treatment and Feedback Generation for an ICALL System in a Second Language Teaching Domain.

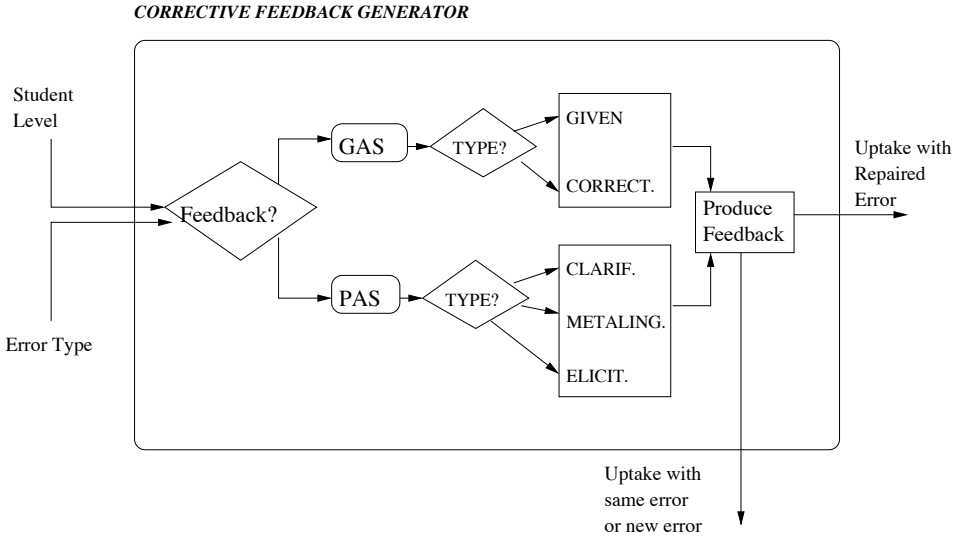

Fig.2. Model of Corrective Feedback Generation.

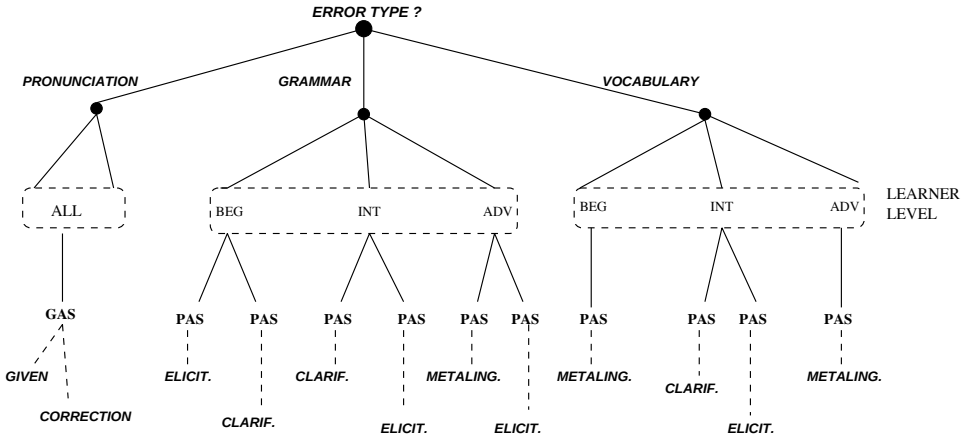

Fig.3. Decision Tree for Feedback Generation after the first error.

```
if error type == grammar
   then if student level == (or beginner intermediate)
           then choose one of the PAS:
              elicitation or clarification-request
           else  ; student level == advanced level
              choose one of the PAS:
                 elicitation or meta-linguistic cue
   else if error type == vocabulary
           then if student level == (or beginner advanced)
                   then choose PAS:  meta-linguistic cue
                   else  ; student-level == intermediate
                      choose one of PAS:
                         clarification-request or elicitation
   else ; error type == pronunciation
      choose one of GAS:
         give answer or explicit correction.
```

This algorithm is graphically depicted in the decision tree of Figure 3.
After the feedback has been generated, the student may produce several types of responses (uptake):

1. An immediate uptake in which the student modifies his/her answer correctly, either by self-repair (if PAS was generated) or by other-repair (if GAS was generated). This indicates that the student has noticed the error and the given assistance, and the correct answer may indicate a first step towards improvement.

2. An uptake which still contains the error. This may occur because the student did not notice the target form provided by the teacher's feedback or the student does not know how to correct the error. In cases such as this, our human teachers either try an alternative feedback strategy or continue the discussion with the next question, an accept turn, or a domain turn. For an ICALL system, we recommend that feedback is selected according the following rule:

```
If this is the second or subsequent corrective feedback
   then if error type == (or grammar vocabulary)
           then choose the most effective PAS strategy
                not already applied
           else ; error type == pronunciation
                ; or all PAS strategies tried
                choose the most effective GAS strategy
                not already applied
```

3. An uptake in which the student repairs the original error, but his/her answer contains another error. In this case, a feedback strategy is selected according to the algorithm for presenting the first corrective feedback move over an error given above.

A remaining issue that must be addressed in any implementation of our model is how the feedback strategies should be realized in natural language and presented to the student. This will depend on aspects of the overall ICALL system, such as whether the student interacts with the system using speech or typing (or a combination), whether the ICALL system is taking a Focus on Forms or Focus on Meaning approach, and so on.[5] We have implemented a simple web-based system that uses a Focus on Forms approach to teaching the Spanish subjunctive. Students are presented with a context in which they have to fill in the appropriate form of a given verb, and receive GAS or PAS feedback (Ferreira-Cabrera, 2003). This system provides one solution to the problem of realizing the feedback strategies.

CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we studied the usage and effectiveness of feedback strategies in Spanish as FL classroom and tutorial interactions, with the aim of providing guidelines for researchers developing feedback strategies for ICALL systems. To this end, we focused on identifying the factors that should be taken into account in order to determine the feedback strategy that is most likely to be effective in any given situation. From these results, we were able to provide an initial model for a feedback component for an ICALL system for Spanish as FL. However, this model is necessarily simplified, and leaves several issues for future research. Although our studies indicate that there is much to learn from examining human-human instructional interactions, we believe that the most fruitful way to proceed in answering many

[5] In the **Focus on Forms** approach, instruction seeks to isolate linguistic forms in order to teach and test them one at a time. In contrast, in the **Focus on Meaning** approach grammar is considered to be best learned incidentally and implicitly through conversation.

of the remaining questions is to build ICALL systems where we can systematically vary the features of interest.

For example, in our study, we found that teachers frequently did not give students the opportunity to answer after a GAS strategy was used, and therefore an important issue for future research relates to wait time (i.e., the length of time that a teacher waits after asking a question before prompting, rephrasing or redirecting the question). Our results indicating that interactive repair rates are substantially higher than overall repair rates, especially in the case of GAS feedback for grammar and vocabulary errors, suggest that significant improvements in instructional effectiveness could be gained by increasing the amount of interaction in the classroom or tutorial setting. One way to increase interaction is to increase wait time. Studies in science education have shown that when wait times are increased (to between three and seven seconds), students respond with more thoughtful answers and that science achievement is increased (Rowe, 1987; Tobin, 1985). In addition, Fox (Fox, 1993) found that when tutors hesitate (i.e., wait) before producing feedback, self-repair increases. However, other studies suggest that increasing wait time may cause students to become more apathetic, or lead to a decrease in achievement, and that this may be caused by waiting too long after low-level questions. These studies suggest that the optimal wait time for a given type of question should be adjusted to the cognitive level of the question, and student's responses should be carefully monitored. Therefore, to determine suitable wait times for questions posed in foreign language teaching at the different levels, we would like to perform studies in which we manipulate this variable to explore how wait time after corrective feedback provision interacts with students' ability to notice the error and self-repair.

Second, in our study, we assessed effectiveness by analysing what happens in the student's next three turns after the teacher draws attention to, corrects, or gives information, or elicits completion in response to the student's incorrect answer. Although the relationship between the frequency of repairs and frequency of each type of corrective feedback gives an indication of the immediate effectiveness of the feedback, this commonly used notion of effectiveness clearly has limitations. Further experiments focusing on how learners respond to the different feedback strategies, and whether they result in changes to their learning in the long term must be performed, and the tendencies found in our studies must be verified in longitudinal experiments. Again, such studies would be facilitated if they were performed in the context of an ICALL system. Indeed, we have performed one longitudinal study with a simple web-based system for teaching the Spanish subjunctive, in which we compared the learning gains (as assessed by pre- and post-test scores) of students who received exclusively GAS feedback to a group of students who received PAS feedback in five sets of exercises administered over a five week period. See (Ferreira-Cabrera, 2003) for more information.

Third, in this study, we have largely focused on the first corrective feedback strategy a teacher provides after a student error. We would like to investigate how FL teachers combine feedback strategies, and whether they tend to use sequences of strategies in a particular way. For example, if a PAS strategy such as elicitation or clarification-request fails to activate previous knowledge, do teachers follow up with a GAS strategy that directly corrects or provides the answer? Knowing how to react if the chosen feedback strategy fails is crucial for the development of robust ICALL systems. Are there scaffolding techniques for certain types of errors or combination of strategies which work this way? To study such sequences will require a larger corpus than the ones used here, and collection and analysis of such

corpora are labor intensive. Alternatively, once we have constructed an ICALL system in which we have implemented a variety of the feedback strategies, we could systematically test sequences which we have observed even in our small corpus.

Overall, our research approach has been enriched by the research in several disciplines, including Second Language Acquisition, Intelligent Tutoring Systems, and Intelligent Computer Assisted Language Learning. These diverse perspectives lead to general questions about how ICALL systems can contribute to alleviating the limitations or disadvantages presented by classroom mode in the treatment of errors, such as giving more opportunity for interaction, prompting student-generated repair. Moreover, the necessity of implementing feedback strategies in ICALL systems can expand our understanding of this key issue and enable us to envisage the kind of contribution that can be useful for ICALL systems, as well as teacher training in the context of foreign language instruction.

References

Aleven, V., Popescu, O., Ogan, A., & Koedinger, K. R. (2003). A formative classroom evaluation of a tutorial dialogue system that supports self-explanation. *Supplemental Proceedings of the 11th International Conference on Artificial Intelligence in Education* (pp. 345–355). Sydney, Australia.

Allen, J. A., & Core, M. (1997). Draft of DAMSL: Dialog act markup in several layers. Available on: *http://www.cs.rochester.edu:80/research/trains/annotation/RevisedManual/RevisedManual.html*.

Allwright, R. (1975). Problems in the Study of the Language Teacher's Treatment of Learner Error. *TESOL'75 New Directions in Second language learning, teaching and bilingual education* (pp. 96–109). Washington, DC.

Allwright, R., & Bailey, K. M. (1991). *Focus on the language classroom*. Cambridge: Cambridge University Press.

Anderson, J., Corbett, A., Koedinger, K., & Pelletier, R. (1995). Cognitive Tutors: Lessons Learned. *The Journal of the Learning Sciences*, 4(2), 167–207.

Bratt, E. O., Schultz, K., Peters, S., Chen, T., & Pon-Barry, H. (2005). Empirical foundations for intelligent coaching systems. *Proceedings of the Interservice/Industry Training, Simulation and Education Conference* (pp. 1–12). Orlando, Florida.

Bull, S. (1997). Promoting effective learning strategy use in CALL. *Computer Assisted Language Learning*, 10(1), 3–39.

Chaudron, C. (1977). A Descriptive Model of Discourse in the Corrective Treatment of Learners' Errors. *Language Learning*, 27, 29–46.

Chaudron, C. (1988). *Second Language Classrooms: Research on Teaching and Learning*. Cambridge University Press.

Chi, M., Bassok, M., & Lewis, M. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, 15, 145–182.

Chi, M. T. H., de Leeuw, N., Chiu, M-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18(3), 439–477.

Chi, M. T. H., Siler, S., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from tutoring. *Cognitive Science*, 25(3), 471–533.

Clark, H. H. & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13(2), 259–294.

Core, M., Moore, J., & Zinn, C. (2002). Tutorial Annotation Scheme. Technical Report September 22, University of Edinburgh.

Criswell, E., Byrnes, H., & Pfister, G. (1991). Intelligent automated strategies of teaching foreign language in context. In M. Swartz & M. Yazdani (Eds.) *Intelligent Tutoring Systems for Foreign Language Learning: The bridge to international communication* (pp. 307–319). Berlin: Springer-Verlag.

Doughty, C. (1994). Focus on form - what is it? *Second Language Research Forum, Montreal*.

Doughty, C. & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.) *Focus on Form in classroom second language acquisition* (pp. 114–138). Cambridge: Cambridge University Press.

Ellis, R. (1997). *The Study of second language acquisition*. Oxford: Oxford University Press.

Fanselow, J. (1977). The treatment of learner error in oral work. *Foreign Language Annals*, 10, 583–593.

Farrar, M. (1990). Discourse and the acquisition of grammatical morphemes. *Journal of Child Language*, 17(3), 607–624.

Ferreira-Cabrera, A. (2003). *Feedback Strategies for Second Language Teaching with implications for Intelligent Tutorial Systems*. PhD thesis, University of Edinburgh, Edinburgh, UK.

Fox, B. (1993). *The Human Tutorial Dialogue Project: Issues in the Design of Instructional Systems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Garrett, N. (1995). ICALL and Second Language Acquisition. In V. Holland, J. Kaplan & M. Sams (Eds.) *Intelligent Language Tutors: Theory shaping technology* (pp. 348–349). Hillsdale, NJ: Lawrence Erlbaum Associates.

Graesser, A., Lu, S., Jackson, G., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M. (2004). Autotutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers*, 36,180–193.

Holley, F., & King, J. (1974). Imitation and Correction in Foreign Language Learning. In J. Schumann & N. Stenson (Eds.) *New frontiers in second language learning* (pp. 81–89). Rowley, MA:Newbury House.

Hume, G., Michael, J., Rovick, A., & Evens, M. (1996). Hinting as a tactic in one-on-one tutoring. *The Journal of the Learning Sciences*, 5(1), 23–47.

Levin, L., & Evans, D. (1995). ALICE-chan: A case study in ICALL theory and practice. In V. Holland, J. Kaplan & M. Sams (Eds.) *Intelligent Language Tutors: Theory shaping technology* (pp. 77–97). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lightbown, P. (1998). The importance of timing in focus on form. In C. Doughty & J. Williams (Eds.) *Focus on Form in classroom second language acquisition* (pp. 177–196). Cambridge: Cambridge University Press.

Lightbown, P. & Spada, N. (1999). *How Languages are Learned*. Oxford: Oxford University Press.

Long, M. (1977). Teacher feedback on learner error: mapping cognitions. In C.E. Snow & C. A. Ferguson (Eds.) *Talking to children: language input and acquisition* (pp. 278–294). Cambridge: Cambridge University Press.

Long, M., Inagaki, S., & Ortega, L. (1997). The Role of Implicit Negative Feedback in SLA: Models and Recasts in Japanese and Spanish. *Modern Language Journal*, 82(3), 357–371.

Lyster, R. (1998a). Negotiation of Form, Recasts, and Explicit Correction in Relation to Error Types and Learner Repair in Immersion Classrooms. *Language Learning*, 48(2):182–218.

Lyster, R. (1998b). Recasts, Repetition, and Ambiguity in L2 Classroom Discourse. *Studies in Second Language Acquisition*, 20, 51–81.

Lyster, R., & Ranta, L. (1997). Corrective Feedback and Learner Uptake: Negotiation of form in Communicative Classrooms. *Studies in Second Language Acquisition*, 19(1), 37–66.

Mackey, A., Gass, S., & McDonough, K. (1990). How do learners perceive interactional feedback? *Studies in Second Language Acquisition*, 22, 471–497.

Merrill, D., Reiser, B., Ranney, M., & Trafton, J. (1992). Effective Tutoring Techniques: Comparison of Human tutors and Intelligent Tutoring Systems. *The Journal of the Learning Sciences*, 2(3), 277–305.

Nagata, N. (1993). Intelligent computer feedback for second language instruction. *Modern Language Journal*, 77(3), 331–339.

Nagata, N. (1995). A study of consciousness-raising by computer: The effect of metalinguistic feedback on second language learning. *Foreign Language Annals*, 28(3), 337–347.

Nagata, N. (1997). The effectiveness of computer-assisted metalinguistic instruction: A case study in japanese. *Foreign Language Annals*, 30(2), 187–199.

Rosé, C. P., Jordan, P., Ringenberg, M., Siler, S., VanLehn, K., & Weinstein, A. (2001). Interactive conceptual tutoring in Atlas-Andes. In J. D. Moore, C. L. Redfield & W. L. Johnson (Eds.) *Artificial Intelligence in Education: AI-ED in the wired and wireless future* (pp. 256–266). Amsterdam: IOS Press.

Rowe, M. (1969). Science, silence and sanctions. *Science and Children*, 6(6), 12–13.

Rowe, M. B. (1987). Wait time: Slowing down may be a way of speeding up. *American Educator*, 11, 38–43.

Sams, M. (1995). Advanced Technologies for Language Learning: The Bridge project. In V. Holland, J. Kaplan & M. Sams (Eds.) *Intelligent Language Tutors: Theory shaping technology* (pp. 7–21). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schmidt, R., & Frota, N. (1986). Developing Basic Conversational Ability in a Second Language: A Case Study of an Adult Learner of Portuguese. In Day, M (Ed.) *Talking to Learn* (pp. 237–326). Rowley, MA: Newbury House.

Seedhouse, P. (1997). The case of missing 'no': The relationship between pedagogy and interaction. *Language Learning*, 47, 547–583.

Sorace, A. (1985). Metalinguistic Knowledge and Language Use in Acquisition-Poor Environments. *Applied Linguistics*, 6(3), 239–254.

Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible output in its developtment. In S. Gass & C. Madden (Eds.) *Input in second language acquisition* (pp. 235–256). Rowley: MA: Newbury House.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlehofer (Eds.) *For H. G. Widdowson: Principles and practice in the study of language* (pp. 125-144). Oxford: Oxford University Press.

Tam, Y. C., Mostow, J., Beck, J., & Banerjee, S. (2003). Training a confidence measure for a reading tutor that listens. *Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech 2003)* (pp. 3161–3164). ISCA Archive.

Tobin, K. (1985). The effect of an extended teacher wait time on science achievement. *Journal of Research in Science Teaching*, 17, 469–475.

Tomlin, R. (1995). Modeling Individual Tutorial Interactions: Theoretical and Empirical Bases of ICALL. In V. Holland, J. Kaplan & M. Sams (Eds.) *Intelligent Language Tutors: theory shaping technology* (pp. 221-241). Hillsdale, NJ: Lawrence Erlbaum Associates.

Van Lier, L. (1988). *The Classroom and the Learner*. London: Longman.

Zinn, C., Moore, J. D., & Core, M. (2005). Intelligent information presentation for tutoring systems. In O. Stock & M. Zancanaro (Eds.) *Multimodal Intelligent Information Presentation*, volume 27 of *Text, Speech and Language Technology* (pp. 227–254). Amsterdam: Kluwer Academic Publishers.