Could You Repeat the Question?

Florence Reeder

George Mason University
freeder@mitre.org

Abstract. The goal of the FLUENT system is to provide authentic language practice for language learners. Misunderstandings in learning systems occur – they are the basis for teaching and tutoring. There is a need, then, for error repair. In starting to build a module to handle language learner errors in a conversational practice system, there are a number of disciplines to consider: language teaching, intelligent tutoring systems, natural language processing and computational linguistics. Correlating the various pieces of research has proven a challenge – because each paradigm utilizes its own analysis formalisms. This paper, then, asks the question: if there is a learner error, what is the appropriate range of responses for misunderstanding? To get an answer to this, we describe our work in progress which is describing error repair for language learning; beginning a taxonomy of repair strategies; determining different analysis strategies and pointing to future work.

Introduction

Our goal is to provide authentic language practice for language learners with a Computer Assisted Language Learning (CALL) system. The developed system, FLUENT (Hamburger, 1995), was initially developed according to Intelligent Tutoring System (ITS) and language pedagogy requirements. We will give a brief introduction to the system which provides the framework and testbed for the work described here. In developing this system, it became desirable to fashion a system which could act as a conversational partner. To be able to build an authentic language partner who responds according to both pedagogical and conversational requirements, we began by looking at classroom interactions and other foreign language learner data. Of particular interest is the handling of errors in interaction. We should design a system that is both conversationally accurate and pedagogically sound for language learning. The process of system design leads us to the following steps: review existing literature on discourse for language learners and in the classroom; select a discourse formalism to model the phenomena we wished to support; analyze classroom and informal interaction data with the formalism; and finally implement appropriate modules for this formalism. This paper, a description of work in progress, describes the system framework in which we will be working; highlights error repair; reviews some previously implemented strategies for handling student misunderstanding; and, finally, explores the problems of trying to reconcile different discourse formalisms and strategies.

FLUENT

FLUENT (Hamburger, 1995) lets a student progress by communicating rather than by explicit study of rules and vocabulary. In Fluent, the student encounters a realistic situation and communicates about it with the system. The communication media are graphical, textual and spoken. Authenticity is achieved by graphically depicting familiar environments, such as a kitchen, living room or office, and performing everyday tasks in these environments. To perform the simple tasks in these microworlds, either the system or the student can pick things up and put them down, turn things on and off, open and close things and so on. The system can move a hand around in the microworld to perform these actions and the student does them using the mouse to control the hand, which acts like a sophisticated cursor.

Language is learned through usage and authentic practice as shown by work on contrasting articles (Schoelles & Hamburger, 1995). Prepositions of location and direction as well as verbs expressing actions

can be learned by a student because the words are coordinated with actual actions done by the student. Second, a student can learn more about words than just their core meaning – context provides subcategorization information for verbs, article usage for nouns and so on. By immersing the student interactively in an ongoing situation and conversation, Fluent fosters a learning style analogous to Total Physical Response, a method with proven success in motivating students to learn linguistic constructs through exposure to variations on similar situations (Asher, 1977).

Fluent draws from the field of Intelligent Tutoring Systems to implement this pedagogy in a systematic way. An ITS typically is divided into functional models which interact to achieve the desired learning effect. A domain knowledge module contains a representation of the target material. Often this knowledge representation is a set of rules, which model the knowledge of an expert in the domain. This material is conveyed to the student through an environment module such as a microworld or simulation. A decision module infers the state of the student's progress and decides which material in the domain module to convey to the student. The inferences are often based on some model of the student's knowledge. These modules all work together to provide learning experience that is at the right level for the student in the context of an appropriate situation. Since FLUENT targets realistic language practice with pedagogical motivation, while maintaining a more traditional ITS approach, work in this vein must draw from multiple areas of research to add new capability. In looking at the handling of student errors, we must describe the error repair context, then look at the issues of finding relevant research from many domains.

Error Repair

In an environment where a user is interacting with a system, errors in understanding will occur. Either the system fails to correctly process an utterance or the user has spoken incorrectly. Sometimes conversants keep quiet and replay the conversation in their head. More frequently we interrupt the flow of conversation and ask for clarification. Repair dialogues capture the human ability to retrace the steps of a conversation and negotiate the proper interpretation of utterances. The goal for language learning environments is the positive handling of these negative situations which is minimally disruptive to the conversational flow.

Initial strategies for Intelligent Tutoring Systems focused on Explanation Generation (e.g., Moore, 1995; Cawsey, 1992; Maybury, 1992). Each follows a slightly different discourse model, but all share many characteristics. All use some form of plan primitives to represent the form of a set of utterances and then a planning system to organize them. The nature of the dialogs tends to fit a more traditional model of interaction: present a problem, ask some questions, carry on a dialog explaining relevant phenomena. This is a more receptive approach than interactive one and difficult in CALL situations where user language skills are weak. Moore (1995) describes limitations of the traditional ITS model including the ability of ITS to account for different learning strategies. The following illustrates how the basic ITS approach could be ineffective for language learners:

> S: The papers has a mistake.
>
> T: In this example, you have used the plural form of **paper(s)** with the singular form of the verb **to have**. Correct usage is to have agreement between forms.

In many current CALL systems, errors are responded to with template-based explanations or by simple right/wrong indicators. Occasionally, a system will try to convey the "expert" knowledge of language to the student (e.g., Kempen & Dijkstra, 1994). Similarly, the system Repairing Errors in Computer Assisted Language Learning (RECALL) diagnoses the student input based on a target language module (Murphy et al., 1997). Other times, systems will point out an error and use template-based explanation of the error (e.g., Schwind, 1994). Most of these try to analyze the type or level of the error through such devices as MAL-rules (e.g., McCoy et al., 1994) and respond appropriately (e.g., Van de Plassche, 1994; Schwind, 1994). Each of these is a valid teaching strategy, but does not reflect the range of pedagogical responses, especially when authentic interaction is a goal. The appropriate handling of an error depends on the goals of the system and the nature of the interaction. If the given utterance is completely unintelligible ("water hand place"), a system may need to respond differently than if the learner merely missed an article ("pick up pot"). Even the detection of errors can depend on the context of an utterance. In the sequence:

Where are the pens located?
The pencil is blue.

the response is grammatically and semantically correct, yet is wrong for the context. The appropriate response depends on the currently employed teaching strategy. In a reflective strategy, the system responds in a way that reflects both recognition of the error and correction of it. ("I know that the stapler is blue, but we were talking about pens.") We are attempting to quantify this, provide a taxonomy and the features of repair that are relevant to different strategies. Two research communities, pedagogy and linguistics, address aspects of the responses, but nothing yet combines the two types of responses.

There is a continuous trade-off between the desire to have students generate exactly right language and the desire to have the system react more realistically. That is, when language learners are faced with real-life situations, they are generally not told that the direct object takes the accusative case as opposed to the dative case. Unless the language is horribly butchered, communication takes place. The language system designer who incorporates dialog-based exercises is faced with the dilemma of determining when, how and what form feedback occurs. In the argument for immediate and complete feedback, one must acknowledge that feedback is essential in the acquisition of skills (Dede & Lewis, 1995). The ability of computers to provide non-threatening, immediate feedback is generally advertised as an advantage of the systems. Usually, a language learner who makes mistakes in a real-life situation can get immediate feedback, although not necessarily the most helpful feedback (i.e., "Why did you call my mother a fish?"). At the same time, continuous tutoring interruption by the system reduces the authenticity of the dialog and can cause anxiety. CALEB (Woolf & Cunningham, 1987) tries to address this, but not in a conversational way.

Current discourse research in error repair seeks to categorize the level of error and determine an appropriate response for each level of error. Levels of errors are determined by the processing layer where failure occurs. For instance, asking to fill the stove with water, while syntactically correct is impossible. This is a knowledge-base level error which may require a different tutoring strategy for repair than incorrect syntactic usage. In an authentic conversational environment, these levels of error repair may be more appropriate than those determined by tutoring strategy and pedagogical needs. At the same time, because this is a learning experience, the tutorial requirements may override authenticity. From a linguistic perspective, McHoul (1990) has analyzed errors and responses in the classroom context and Mostow and Aist (1998) have looked at the disruptive nature of correction for reading exercises. Here are possible reactions to the same error based on these two, sometimes conflicting, needs.

S: Pick up cup.
T: I am picking up **the** cup.
Reaction 1: Understanding Incorrect Utterance, Responding with Correct Answer

S: Pick up cup
T: That is incorrect. You must say "the" before talking about an object.
Reaction 2: Template Explanation for Incorrect Utterance

S: Pick up cup
T: You mean, pick up the cup.
S: Yes, pick up the cup.
Reaction 3: Correction with Verification

Analysis of Classroom and Learner Interactions

Because of the need to design a system that is both pedagogically sound and conversationally accurate, we begin by reviewing literature in these areas. At this point, the confusion created by a cross-disciplinary effort appears. Literature describing classroom interactions and error repair with language learners can be found in psycholinguistics, pedagogy, artificial intelligence (AI) and natural language processing (NLP). This means differing models and descriptions must be reconciled if our system is to benefit from work previously accomplished in these fields. The formalisms we will examine here are representative of the

communities: Conversational Analysis (CA); the DISCOUNT scheme (Pilkington, et al., 1999); and the DAMSL scheme (Allen & Core, 1997). This is not an exhaustive list of dialog annotation and discourse analysis paradigms – more need to be considered including TRINDI (Bohlin, et al., 1999) and others. The field of tutorial dialogue is a burgeoning one. So, we need to incorporate recent discourse work by Clark (1996) and his students to ensure that pedagogical interaction is effectively captured.

Conversational Analysis (CA) and the work of Sinclair and Coulthard (1985) describe interactions by a sequence of speech-act pairs such as greet/acknowledge. The appropriate exchange sequences are dependent on the context of their use. An example of this is the Initiate/Response/Follow-Up sequence for error correction. Implementing work based on this paradigm and the research which uses it is difficult because while the technique is descriptive, it is not quite sufficient for implementation in software. More recent work in analyzing classroom and tutorial interaction (e.g., Clark, 1996; Craig, et al., 2000) will have to be examined before a final scheme can be determined.[1]

The DISCOUNT scheme has the advantage of being designed to support discourse analysis for learning situations. Since it has been developed by NLP researchers, it should translate into a computational framework readily. It is, to some extent, built upon the ideas of CA – particularly the exchange structure analysis. Yet, it differs from CA sufficiently to make an analysis which synthesizes work from both communities difficult. Since it is a multi-layer approach, dialog tagging will be more complex.

Finally, the DAMSL scheme is a generalized dialog markup language originally developed as part of the Discourse Research Initiative (DRI) which was seeking to standardize discourse tagging and cataloging. The scheme is very general and represents an NLP view of interaction. It presents, minimally, the kinds of interactions found when a human is interacting with a computer for the purpose of solving a common goal. Although the formalism more readily corresponds to computing needs, it is not as directly applicable to the language teaching experience as one might hope. The next step for selecting a formalism and utilizing it is to take a sample set of dialogs, both of classroom and language learner interactions and apply these formalisms to them. This exercise will readily point out the positives and deficiencies of all formalisms and should indicate which is the most appropriate for use in our system.

Designing Repair Strategies

In designing a system to support language teaching through simulation and interaction, we want to be responsive to the pedagogical, design and architectural issues. Additionally, it is desirable to build on existing work instead of re-inventing interaction software for each new tutoring domain. Part of the difficulty of developing systems is being able to utilize and synthesize the wide body of research accomplished in multiple fields. For language learning, the fields of pedagogy, language acquisition, linguistics and natural language processing (NLP) contain relevant information for system development. Each of these fields has its own research methodologies, paradigms and descriptions of the activities and requirements of a system to fit into teaching. Because of these differing views, system designers must correlate different research in the fields. One aspect of this is the differentiation between feedback, explanation and error repair. Each of these techniques can be used for answering a student's misconception. An initial survey of the fields (ITS, NLP, SLA) points out three techniques which can overlap – from pedagogy (feedback); from linguistics (error repair) and from NLP (explanation generation). This section looks at their similarities and differences with a view towards developing a framework for ITS system responses.

This stage of development is difficult because of various analysis and theoretical formalisms. Traditional discourse modules treat error repair according to the component of the system which makes the error. But, CALL systems treat the error according to the type of error. Pedagogists echo this approach – and in fact argue that different error types merit different types of repair. Yet, a conversational practice system may need a range of repairs. A good system design could allow us to test out the various repair types. There is a need, then, to look at the range of responses to error situations. The range of error situation responses has never been clearly elucidated and the distinctions really get at the heart of teaching. The first step is to define this problem – what is error repair, what is teaching and where do the various

[1] The author would like to thank the reviewers for the helpful pointers to this material.

existing techniques fit into the hierarchy. The second step is to then identify, from a system design perspective, the appropriate structures to support what is needed in an ITS system.

Second language acquisition researchers have a definite notion of error repair in the classroom (e.g., Lyster, 1998). Rather than being a negotiation of meaning, error repair is the correction of, typically, form errors in language learning exercises. On the other hand, in traditional ITS literature and NLP literature, error repair is treated as the negotiation of understanding (e.g., Purlis & Purlang, 1996; Dybkjær et al., 1996). Finally, a body of work that the team is beginning to work with looks at the tutorial interaction on general subjects and with computers as well (Cox, et al, 1999). Each of these looks at the form of the error – is it a question or a mistake – and the form of the repair – an answer; a correction or other. Integrating these is a big challenge – particularly in line with building a useful, system. We will be continuing the analysis and taxonomy work before selecting one particular avenue of research.

Conclusion

Our goal, then, is to determine from classroom and interaction data, what are the possible responses to misunderstanding. By the workshop, we expect preliminary results to indicate the appropriate discourse formalisms. However, feedback from reviewers has pointed to a new body of research, particularly in the tutorial dialogue domain that has increased the challenge for a comprehensive picture of the field. In addition, we can then use these results to begin to describe the nature of the teaching interaction – by defining feedback, error repair and pedagogical instruction and their relationship to each other.

References

Asher, A. (1977). Learning Another Language Through Actions: The Complete Teachers Guidebook. Los Gatos, CA: Sky Oaks Productions.

Allen, J. & Core, M. 1997. *Draft of DAMSL: Dialog Act Markup in Several Layers*. Technical Report from Discourse Research Initiative (DRI) http://www.cs.rochester.edu:80/research/trains/annotation/RevisedManual/RevisedManual.html

Bohlin, P., Bos, J., Larsson, S., Lewin, I., Matheson, C., & Milward, D. 1999. Survey of Existing Interactive Systems. Task-Oriented Instructional Dialogue, TRINDI Technical Report, LE4-8314.

Cawsey, A. 1992. *Explanation and Interaction*. Cambridge, Mass: MIT Press.

Clark, H. H. 1996. *Using Language*. Cambridge University Press.

Coulthard, M. 1985. *An Introduction to Discourse Analysis*, 2nd ed. Longman Press

Cox, R., McKendree, J., Tobin, R., Lee, J. & Mayes, T. 1999. Vicarious learning from dialogue and discourse: A controlled comparison. *Instructional Science* 27: 431-458.

Craig, S., Gholson, B., Ventura, M., Graesser, A. & The Tutoring Research Group. 2000. Overhearing Dialogues and Monologues in Virtual Tutoring Sessions: Effects on Questioning and Vicarious Learning. *International Journal of Artificial Intelligence in Education* (to appear).

Dede, C. & Lewis, M. 1995. Assessment of Emerging Educational Technologies That Might Assist and Enhance School-to-Work Transitions. Report for Office of Technology Assessment.

Dybkjær, L., Bernsen, N., & Dybkjær, H. 1996. Reducing Miscommunication in Spoken Human-Machine Dialogue. In S. McRoy, ed., *Workshop on Detecting, Repairing, and Preventing Human-Machine Miscommunication, AAAI-96, Thirteenth National Conference on Artificial Intelligence*, Workshop Working Notes. pp. 29 – 36.

Hamburger, H. 1995. Tutorial Tools for Language Learning by Two-Medium Dialog. In V. M. Holland, J. D. Kaplan, M. R. Sams (Eds.), *Intelligent Language Tutors: Theory Shaping Technology*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kempen, G., & A. Dijkstra. (1994) Toward an Integrated System for Grammar, Spelling and Writing Instruction. In L. Appelo & F.M.G. de Jong (eds.), *Computer-Assisted Language Learning, Proceedings of the Seventh Twente Workshop on Language Technology*. Enschede: University of Twente.

Lyster, R. 1998. Negotiation of Form, Recasts, and Explicit Correction in Relation to Error Types and Learner Repair in Immersion Classrooms. *Language Learning.* 48(2), 183-218.

Maybury, M. 1992. Communicative Acts for Explanation Generation. *Int. Journal of Man-Machine Studies*, 37, 2, 135-172.

McCoy, K., Pennington, C. & Suri, L. 1996. English Error Correction: A Syntactic User Model Based on Principled "Mal-Rule" Scoring. In *Proceedings of the Fifth International Conference on User Modeling*. Kailua-Kona, Hawaii.

McHoul, A. 1990. The organization of repair in classroom talk. *Language in Society*, 19, 3, 349-377.

Moore, J. 1995. *Participating In Explanatory Dialogs.* Cambridge, Mass.: MIT Press.

Mostow, J. & Aist, G. 1998. Adapting Human Tutorial Interventions for a Reading Tutor That Listens: Using Continuous Speech Recognition in Interactive Educational Multimedia. In Cameron, K. (Ed.), *Multimedia CALL: Theory and Practice*, Exeter: Elm Bank Publications.

Murphy, M., Kruger, A & Grieszl, A. 1997. RECALL: Towards a Learner Centered Approach to Language Learning. In du Boulay, B. & Mizoguchi (Eds.), *Proceedings of Artificial Intelligence in Education*, IOS Press.

Pilkington, R., T. Treasure-Jones & C. Kneser. 1999. The Tutor's Role: An Investigation of Participant Roles in CMC Seminars Using Exchange Structure Analysis. In R. Pilkington, ed., *Analysing Educational Dialogue Interaction: Towards Models that Support Learning.* Proceedings of Workshop at the 9th International Conference on Artificial Intelligence in Education (AI-ED 99). Le Mans, France, p. 44 – 53.

Schwind, C. 1994. Error Analysis and Explanation in Knowledge-Based Language Tutoring. In L. Appelo & F.M.G. de Jong (eds.), *Computer-Assisted Language Learning, Proceedings of the Seventh Twente Workshop on Language Technology*. Enschede: University of Twente.

Schoelles, M. and Hamburger, H. 1995. Cognitive tools for language pedagogy. In *Computer-Assisted Language Learning*. The Netherlands: Swets & Zeitlinger.

Van de Plassche, J., E. Bos & E. Jongen-Junner. 1994. Verbarium and Substantarium: Two CALL Programs for Latin. In L. Appelo & F.M.G. de Jong (eds.), *Computer-Assisted Language Learning, Proceedings of the Seventh Twente Workshop on Language Technology.* Enschede: University of Twente.

Woolf, B., & Cunningham, P. 1987. Multiple Knowledge Sources in Intelligent Teaching Systems. *IEEE Expert*. p. 41 - 54.