Could You Repeat the Question?

Florence Reeder

George Mason University / MITRE Corporation
1820 Dolley Madison Blvd.
McLean, VA 22102
freeder@mitre.org

Abstract

The goal of the FLUENT system is to provide authentic language practice for language learners. In learning systems, misunderstandings occur – they are the basis for teaching and tutoring. There is the need, then, for error repair. This paper asks the question: if there is a learner error, what is the appropriate range of responses to resolve the misunderstanding? In starting to build such a module, a number of disciplines come into play: language teaching, intelligent tutoring systems, natural language processing and computational linguistics. We will touch on the difficulty in synthesizing each of these domains as we begin to analyze the handling of errors in a language learning environment.

Introduction

Our goal is to provide authentic language practice for language learners with a Computer-Assisted Language Learning (CALL) system. To be able to build an authentic language partner who responds 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: reviewing existing literature on discourse for language learners; select a discourse formalism to annotate the interactions; choose from the annotated interactions error situations to handle; and implement appropriate modules for this formalism. This paper, a description of work in progress, highlights error repair; reviews some previously implemented strategies for handling student misunderstanding; and, finally, explores the problems of selecting a discourse modeling formalism.

Error Repair

In human-human communication, misunderstandings are bound to happen. In response to these, conversants sometimes keep quiet and reply the conversation in their head. More frequently, they 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. 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. It is the goal of system designers to detect the misunderstanding and initiate an appropriate repair strategy to effectively accomplish the given task.

In most teaching and tutoring situations, misunderstandings represent an opportunity to effectively educate the primary holder of misunderstandings – the student. Past Intelligent Tutoring Systems (ITS) utilized these traditional teaching strategies (e.g., Polson and Richardson 1988), relying on techniques such as Explanation Generation (e.g., Moore 1995). These types of systems use plan primitives to represent the form of a set of utterances and then a planning system to organize them. The nature of the dialogues tends to fit a more traditional model of interaction: present a problem, ask some questions, carry on a dialogue explaining relevant phenomena. Recent work (Pilkington 1999) highlights an effort to better reflect the wide variety of teaching and repair strategies in tutoring systems.

Second language acquisition represents a unique domain for educational interaction. The domain of study and the medium of communication is frequently the same – that is, in the communicative classroom based on the pedagogical principle of Total Physical Response (Asher 1977) all interaction is in the language being learned. Authentic language practice is the continuous goal of the classroom, sometimes limiting the range of responses to misunderstandings. Additionally, the elicitation of knowledge from the students can be challenging. If the goal is to teach prepositions, for instance, setting up a "natural" interaction to elicit them is difficult.

As a result of this domain, the ITS that supports language learning represents a different kind of challenge for modeling tutorial dialogue. The goal of a Computer-Assisted Language Learning (CALL) system is to present authentic language practice while providing the correct educational supports. There is a continuous trade-off

between the desire to have the student generate exactly the 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. Authentic human-human repair is negotiated according to traditional, possibly simplified, strategies. The system must be able to support both the authentic repair strategies while maintaining the pedagogical goals of teaching language effectively. We will now look at traditional handling of errors by CALL systems before turning to the desired goals in a CALL language practice system.

CALL Error Handling

Two parts of responding to CALL errors exist: diagnosing the errors and selecting a response. Error diagnosis for CALL is more complex than identifying that an error occurred (Holland et al. 1995). For instance, in subject-verb agreement errors, the system must decide if the student misunderstands either subject-verb agreement or the morphology of the incorrect element. Error detection and diagnosis is accomplished in the natural language understanding (NLU) component of the system (e.g., Heift 1998). McCoy et al. (1994) utilize parsing strategies with MAL-rules to analyze the potential causes of the error. In another variation of this, the RECALL (Repairing Errors in Computer-Assisted Language Learning) system diagnoses the student input based on a target language module (Murphy et al. 1997). Michaud and McCoy (2000) present a complex error diagnosis component which holds the promise of accurate and reasonable error disambiguation since it is based on a model of the stages of language acquisition.

Regardless of the method of diagnosis and detection, errors typically are responded to with template-based explanations or by simple right/wrong indicators (e.g., Michaud and McCoy 1999; Schwind 1994; Van de Plassche et al, 1994). The CALEB system (Woolf and Cunningham 1987) simply points out the error location without explanation. Occasionally, a system will try to convey the "expert" knowledge of language to the student (e.g., Kempen and Dijkstra 1994).

Each of these represents a valid, traditional ITS teaching strategy. Yet they do not reflect the possible range of pedagogical responses, particularly in a communicative classroom where authentic interaction is a goal. The appropriate handling of an error depends on both the goals of the system pedagogically and the nature of the interaction conversationally. 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 pan**. Consider 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 error handling depends on the currently employed teaching strategy. In a reflective strategy, the system responds in a way that shows recognition of the error and correction of it, **I know the stapler is blue, but we were talking about pens**. In a purely authentic language practice session, the system may react to an error as a partner with immediate, but not necessarily instructive, feedback (i.e., **Why did you call my mother a fish?**). Having described error repair as it has traditionally been handled by CALL, we will now focus on the design of repair strategies which can handle both authentic language needs with effective pedagogical techniques.

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 CALL systems, the fields of pedagogy, second language acquisition, linguistics and 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 various research elements from these fields.

One area where the fields differ 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 NLP (error repair) and from ITS (explanation generation). This section looks at their similarities and differences with a view towards developing a framework for ITS system responses.

From a pedagogical and linguistic perspective (Lyster 1998; Seedhouse, 1999) analyze classroom interactions and base recommendations for error repair on the type of error. For instance, in Lyster's work the classroom learner errors are catalogued as one of grammatical, lexical, phonological or L1 usage. These then are matched against responses which include negotiation of form, recasts or explicit correction. The negotiation of form is further broken down into elicitation, metalinguistic clues, clarification requests or repetition of error with emphasis. This work attempts to demonstrate which of the repair strategies is most effective for SLA. Seedhouse similarly categorizes errors, although his analysis argues that the best reply to errors is explicit correction and continuation of the conversation.

Current dialogue management research in error repair seeks to categorize the level of error and determine an appropriate response for each level of error (e.g., LuperFoy and Duff 1996). 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. The repair strategy characterized in these types of systems is similar to the pedagogical analysis in that it determines levels of processing for attributing errors. On the other hand, it shows a wider range of possible error sources while not addressing the finer grained distinctions which may be necessary for effective language teaching.

In an authentic conversational environment, the 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. Table 1 shows 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**

Table 1 Responses for an Error

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

Analysis of 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. 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) and the DISCOUNT scheme (Pilkington, et al., 1999). This is not an exhaustive list of dialog annotation and discourse analysis paradigms – more need to be considered including TRINDI (Bohlin, et al., 1999). Space does not permit a complete assessment of all possibilities at this time.

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.

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 synthesizing work from both communities difficult. Since it is a multi-layer approach, dialog tagging will be more complex. Additionally, work will be necessary to determine if it is sufficient to support the fine-grained distinctions necessary for language learning and authentic language practice.

Conclusion

As stated previously, this effort represents very preliminary work in building error repair dialogues for a tutoring system. We welcome the opportunity to learn from others during this workshop.

Acknowledgments. The views expressed in this paper are those of the author and do not reflect the policy of the MITRE Corporation. Many thanks to the anonymous reviewers for their astute and helpful comments.

References

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

Bohlin, P.; Bos, J.; Larsson, S.; Lewin, I.; Matheson, C.; and Milward, D. 1999. Survey of Existing Interactive

Systems. Task-Oriented Instructional Dialogue, TRINDI Technical Report, LE4-8314.

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

Heift, G. 1998. Designed Intelligence: A Language Teacher Model. Unpublished Ph.D. thesis. Simon Fraser University.

Holland, V. M.; Kaplan, J. D.; and Sams, M. R. (eds.). 1995. *Intelligent Language Tutors: Theory Shaping Technology*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kempen, G.; and Dijkstra, A. 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.

LuperFoy, S.; and Duff, D. 1996. Disco: A Four-Step Dialogue Recovery Program. 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. 73-76.

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.

McCoy, K.; Pennington, C.; and 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.

Michaud, L.; and McCoy, K. 1999. Modeling User Language Proficiency in a Writing Tutor for Deaf Learners of English. In M. Olsen, (ed.), Computer-Mediated Language Assessment and Evaluation in Natural Language Processing, Proceedings of a Symposium by ACL/IALL. University of Maryland, p. 47-54

Michaud, L.; and McCoy, K. 2000. Supporting Intelligent Tutoring in CALL by Modeling the User's Grammar. In *Proceedings of the 13th Annual International Florida Artificial Intelligence Research Symposium*, Orlando, Florida, p. 50-54, May 22-24, 2000.

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

Mostow, J.; and 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.; Krüger, A.; and 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. (ed.). 1999. 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.

Pilkington, R.; Treasure-Jones, T; and Kneser, C. 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.

Polson, M.; and Richardson, J. J. (eds.) 1988. *Foundations of Intelligent Tutoring Systems*. New Jersey: Lawrence Erlbaum Associates

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.

Seedhouse, P. 1999. The Relationship Between Context and the Organisation of Repair in the L2 Classroom. *International Review of Applied Linguistics in Language Teaching*. 37(1): 59-80.

Van de Plassche, J., Bos, E. and Jongen-Junner, E. 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.; and Cunningham, P. 1987. Multiple Knowledge Sources in Intelligent Teaching Systems. *IEEE Expert*. p. 41 - 54.