Modeling User Language Proficiency in a Writing Tutor for Deaf Learners of English

Lisa N. Michaud and Kathleen F. McCoy
{michaud,mccoy}@cis.udel.edu
Computer and Information Sciences Department
University of Delaware Newark, DE 19716

Abstract

In this paper we discuss a proposed user knowledge modeling architecture for the ICICLE system, a language tutoring application for deaf learners of written English. The model will represent the language proficiency of the user and is designed to be referenced during both writing analysis and feedback production. We motivate our model design by citing relevant research on second language and cognitive skill acquisition, and briefly discuss preliminary empirical evidence supporting the design. We conclude by showing how our design can provide a rich and robust information base to a language assessment / correction application by modeling user proficiency at a high level of granularity and specificity.

1 Introduction

In order for any human language tutor to be effective, he or she must have an accurate picture of the student's language acquisition status. This "picture" is used for selecting target features for tutoring and for shaping and tailoring the tutorial instruction. Automated tutoring systems emulate this desirable practice by constructing and maintaining a model of the user's knowledge, consulted at many levels of the tutorial production process. In this paper we will discuss the proposed knowledge modeling architecture of ICICLE (*Interactive Computer Identification and Correction of Language Errors*), a system under development (McCoy and Masterman (Michaud), 1997) whose goal is to provide deaf students with constructive tutoring on their written English.

The target learner group for ICICLE is native or near-native users of American Sign Language (ASL). This population poses unique challenges for a writing instruction program: their writing contains many errors which are not made by native users of English, and students vary widely across levels of language ability, with some college-age writers producing near-native-like English and others struggling with grammatical basics. Because of these characteristics of the learner population, it is integral to ICICLE's goal of user-tailored instruction that it account for user differences so that the instruction it provides is appropriate for a learner at any level. Since ASL is a distinct and vastly different language from English (Baker and Cokely, 1980), we view the acquisition of written English skills to be a task in second language acquisition for these learners (Michaud and McCoy, 1998). We are therefore proposing a user model design which incorporates a representation of the language acquisition process, and we have based our design upon current research in language acquisition and in the acquisition of cognitive skills. ICICLE will consult this model to obtain specific information about the user's current language knowledge, as well as about what knowledge is likely to be learnable by the user at the current time.

In the following sections, we overview the ICICLE system architecture in order to explain how the user model will be utilized in system operation; we then discuss our design for modeling the second language acquisition process, and overview issues involved in implementing the model within our system.

2 ICICLE Overview

ICICLE interacts with its user through a cycle of user input and system response. The user begins this cycle by supplying the system with a multi-sentential piece of writing for analysis. An error identification component uses an English grammar augmented with rules to

cover grammatical errors commonly produced by our learner population to process the user's writing, tagging the grammatical errors it finds (Suri and McCoy, 1993; Schneider and McCoy, 1998). In the current implementation of ICICLE, the identified errors are highlighted in a window-based interface using colors which indicate the class of error; for example, all subject-verb agreement errors are highlighted in blue. The user may then explore particular sentences containing errors by clicking on them, causing an editing window to appear with a simple one-sentence "canned" response explaining the error. The user may then edit the sentence, have it reanalyzed by the system, and paste the results back into the original text.

In the completed system, errors identified by the system will be passed to a response generation component which will construct one or more natural language explanations to discuss the nature of the error(s) with the student, who will then be prompted to make corrections and request a new analysis as in the current implementation.

Both of the active processes in the ICICLE architecture (error identification and response generation) will access the user language acquisition model, which is currently under development. The error identification module will use the model to determine between multiple interpretations of a sentence which may correspond to different perceived errors in the text (McCoy et al., 1996). The essence of this determination is discerning the cause of the error. For instance, if the phrase "My brother like to go..."[1] has occurred in the writing of a student, there are several possible situations that could have led to this mistake: the student could be entirely unaware of the English rule for subject/verb agreement; the student could know of the rule, but has applied it incorrectly; or the student has simply mistyped. To determine which of these possibilities is correct, it is necessary for the error analysis component to have at its disposal a model of the student's language knowledge which indicates his or her mastery of such language features as the concept of subject/verb agreement. (In the present system the choice between multiple interpreta-

---

[1]This example has been taken from our corpus of deaf writing samples.

tions is not yet made on a principled basis.)

We also wish for ICICLE to give instruction only on those language features which are at the user's current level of acquisition; errors on features above this level are likely to be beyond the user's understanding, while errors on features which are well-established are likely to be simple mistakes which do not require instruction. The user model will therefore be consulted at the point where the error identification module passes the list of errors to the response generation module, trimming off those errors outside the current level. The motivation for this action is discussed further in Section 3.

Lastly, the user model will be consulted during the planning of the system response. In order to structure explanations of a given language feature, the text planner needs to know the user's depth of related knowledge, including whether or not the user knows the concepts which are mentioned in the explanation. In the final stages of response generation, the surface generator will also need to consult the model of acquisition in order to determine which grammatical constructs are known and thus understandable to the user, and which he or she may obtain the most benefit from viewing as positive examples.

We have therefore established that a user model which contains a representation of second language proficiency, specific to the detail of individual language features and the user's mastery of them, is essential to the envisioned operation of the ICICLE system. We will now discuss our proposed design for this model and overview some of the issues we face in implementing this design.

3 Modeling Theories of Cognitive Skill and Second Language Acquisition

We see our model as representing the user's location along the path toward acquiring written English as a second language. To design this model, we have looked into the interlanguage theory of second language acquisition (Selinker, 1972; Ellis, 1994; Cook, 1993). In this theory, a learner generates in the L2 using a grammar which is his or her hypothesis of the target language. The learner systematically revises this internal representation as the language is ac-

quired. We espouse the view that the initial interlanguage model formed by the learner is based largely on knowledge from his or her native language (Schwartz, 1998; Schwartz and Sprouse, 1996), and that as the learner progresses, more of the interlanguage is a correct model of the target language, and less reflects the native language or incorrect assumptions. As the interlanguage grammar progresses toward a target-like form, portions of the grammar become the focus of hypothesis testing and thus are somewhat in "flux" between an incorrect form and a target-like form. Each "revision" of the interlanguage results in an increase in target-like features in the grammar. It is our goal to create a model whose contents will reveal the status of these features as a snapshot of the learner's current interlanguage state.

Research in second language acquisition and education indicates that as a learner masters a subject, there is always some subset of that material that is currently within his or her grasp to acquire. Intuitively, it is this area that he or she is currently in the process of learning. This subset has been termed the Zone of Proximal Development (ZPD) (Vygotsky, 1986). The general idea has been applied to assessment and writing instruction by (Rueda, 1990), and to second language acquisition by (Krashen, 1981). In our model, the ZPD corresponds to the portion of the interlanguage currently "in flux" and in the process of making a transition to the target grammar. The identification of the ZPD for a given second language learner would be an ideal indication of the next language features he or she will acquire, or those features on which instruction would be most beneficial because they are neither well-established nor beyond his or her ability to learn at this time.

The goal of identifying the ZPD is aided by the suggestion made by other researchers in first and second language acquisition that the language errors committed by a learner systematically change over time (Dulay and Burt, 1974), and furthermore that there may be a specific sequence of acquisition a learner follows in acquiring language features that may be relatively fixed regardless of the native language (Krashen, 1981; Ingram, 1989; Dulay and Burt, 1975; Bailey et al., 1974). If the learner is following this sequence, identifying the current

stage of the process would have implications for which features have been learned, and which are soon to be acquired.

In order to be able to determine the learner's placement in an order of acquisition of morphosyntactic features, we will examine the user's performance in the target language, and compare it against what we know of the performance of learned cognitive skills. (O'Malley and Chamot, 1990) list the stages *cognitive, associative,* and *autonomous* to explain the progression of a learner through levels of competency. We expect the "in flux" portion of our learner's interlanguage to go through similar stages. At the initial cognitive stage, a language feature has just entered the learner's ZPD. At this point, the learner is aware of the features but the knowledge of how to use them is impoverished or incomplete, so he or she is incapable of performance with consistent skill. This knowledge is thus termed "declarative." At the next stage, the associative stage, the errors in the original declarative representation are systematically deleted while the learner improves his or her understanding. The declarative knowledge develops into "procedural" form, or a form which can be used to successfully utilize the feature. At the final autonomous stage, the performance is fine-tuned, and the skill becomes virtually automatic while errors disappear. This functional knowledge distinction is also similar to that represented in the user model of TAILOR, another tutoring system (Paris, 1987), and is supported by psychological studies such as (Chi et al., 1981).

The fact that declarative knowledge is shallow and results in the production of errors ties these views into the ZPD theory, where the Zone is the area in which one should expect the most errors to occur (Vygotsky, 1986). To apply these theories to second language performance, the errors produced by a learner should predominantly represent the morphosyntactic features in his or her ZPD. Features that have been acquired previously should occur without significant variation or error, and features beyond the ZPD should be absent from his or her language production because they are beyond the learner's knowledge.

We will represent these observations and theories in an overlay model which is an instan-

tiation of a knowledge base of morphosyntactic features, tagged to indicate each feature's placement within a given user's knowledge - "acquired" or "ZPD," depending on the user's performance on each feature. In the next section, we address how we expect to assign these tags in the initial state of the model. After initiation, it would be our expectation that over time those features indicated as being part of the ZPD would be tagged as "acquired" once they are used with consistent correctness, and features that had no tags previously (because they were absent in the learner's language production) would move into the ZPD once the learner is ready to begin acquiring them. The "feature" units in this knowledge base will be those features represented by the augmented grammar which ICICLE uses to parse its input, since that is the granularity of its error analysis capacity and of the feedback the system presents.

This design answers the needs of both active modules of the ICICLE architecture. The error identification phase could use it when selecting a parse for a given portion of text. Because of the relationship between the granularity of the model and the grammar, the action of indicating the ZPD in the model could be mirrored in the grammar, with special notation given to those grammatical rules covering ZPD concepts. The parser can assume that structures tagged as "acquired" in the model representing this user will be used correctly with consistency, while those within the ZPD are most likely to occur with error, and those which are beyond the user's knowledge will be absent from his or her writing. When choosing a parse, the system should favor one using "correct" English grammar rules from the "acquired" range, and "incorrect" rules from the ZPD range. Thus the correct parse and source of error can be determined by comparing the possibilities against what constructions the user is expected to use correctly or incorrectly according to the model.

A model of this type would also provide vital information needed for transforming a list of errors into the tutorial response. Instruction and corrective feedback on aspects of the knowledge within the ZPD may be beneficial, while instruction dealing with that outside of the Zone is likely to be ineffective or even detrimental. Tutoring on material outside the ZPD which has
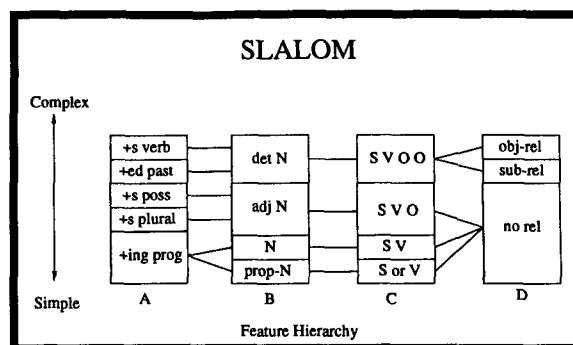

Figure 1: SLALOM: Steps of Language Acquisition in a Layered Organization Model.

already been mastered by the student is likely to bore them; tutoring on material beyond the grasp of the student at this time is likely to produce confusion or frustration. When passing the error list to the response module, the error identification module can use the user's placement in the model to prune the errors so that the tutorial responses are focused only on those errors at the user's current level of language acquisition.

The actual construction of the system response can also reference this model, using it to determine the user's depth of knowledge on the features being discussed so that appropriate background information and definitions of terms being used can be provided. The full interaction between the text planner and the user model is a topic of current exploration.

In formalizing our user model design, we therefore need to capture three aspects of language competence: the past, the present, and the future. The model must be able to indicate which features of language the user has already mastered, those features he is presently attempting to acquire, and those features that are above his current level. The next section discusses how we propose to structure this information in the model, and overviews our approach for building and maintaining it.

4 SLALOM: A Proposed Model Architecture

Our proposed architecture for capturing the theories expressed in the previous section is SLALOM (*Steps of Language Acquisition in a Layered Organization Model*). A very simplified representation of SLALOM can be found in Fig-

ure 1. SLALOM captures the stereotypic linear order of acquisition within certain categories of morphological and/or syntactic features of language, such as NP or relative clause formation. Within a category, depicted as a stack of features in the figure, a given morphosyntactic feature is expected to be acquired subsequent to those below it, and prior to those above it. Lateral connections between the categories indicate features which we expect to be acquired concurrently. As mentioned in the previous section, an instantiation of this model would represent a given user by tagging the features as acquired or within the ZPD according to observations of the user's language performance on texts analyzed by the system. Once such observations have been noted, inferring additional information about non-tagged elements would be possible through exploiting the lateral connections to infer a concurrent relationship or exploiting the orders within hierarchies to infer whether a feature is likely to have been acquired by this user. The explicitly-marked tags may be revised over time as the learner's proficiency develops, with those features tagged as within the ZPD moving to acquired status, and new features from the not-yet-acquired (untagged) area moving into the ZPD. Because the SLALOM architecture represents an expected order of acquisition, the likely path of the ZPD would be to move "up" in the stacks.

Since previous studies to identify orders of language feature acquisition such as (Bailey et al., 1974) and (Dulay and Burt, 1975) focus primarily on morpheme usage and not the higher-order grammatical constructs we also need for our system, identifying a morphosyntactic order of acquisition is an area of current work in which we have preliminary empirical results. A statistical analysis of a corpus of 101 writing samples from deaf students[2] has revealed groupings of morphosyntactic errors exhibiting apparent correlation to a general (and currently subjective) estimate of writing proficiency. This indicates a relationship between the proficiency of the learner and the errors he or she commits. These findings are informal and we are planning further exploration with a larger corpus and more objective judging, but they indicate a

---

[2]These were performed by undergraduate assistant Litza Stark.

basis for the ordered organization of SLALOM. Additional analysis has shown statistically significant relationships of co-occurrence between certain errors among learners of similar proficiency levels, supporting the existence of lateral relationships of concurrent development. This work is ongoing and will hopefully yield more concrete results soon.

5 Implementation Issues

In this section, we address some of the issues involved in making use of a SLALOM-based model within our evaluation/feedback application, including: establishing a user's placement within the model; referencing the model once it has been established; and updating the model over time.

5.1 Establishing and Referencing the Model

Most natural language systems which reference a user model are more concerned with using the model than building it or updating it over time (Ringle and Bruce, 1981; Woolf, 1984; Paris, 1987; Moore and Paris, 1992; Carenini and Moore, 1993). However, since ICICLE's user model needs to capture fine details of user knowledge derived from individual user performance, and ICICLE will be used by a learner over time and across the development of new skills, the system must concern itself with both establishing and updating its model of user knowledge.

We wish to base ICICLE's techniques for building and updating an instantiation of its user model on the sources of user information proposed by (Wahlster and Kobsa, 1986), including:

- *Initial individual models* stored from previous sessions

- *Assumptions based on user input* which provide explicit (direct) or implicit (indirect) inferences

- *Default assumptions* from stereotypic information.

The initiation of the user model in a given session will depend on whether a user has accessed the system in the past; the models of previous users will, of course, be restored from earlier

51

sessions. New users will require the system to initialize the model according to the input they provide in the first sample of writing they enter[3].

In the list of information sources above, we mention both explicit and implicit information provided to the model by user input. In the direct sense, a user's writing is a uniquely rich source of language proficiency information. In comparison to the techniques other systems use to determine user knowledge such as polling, where one question is only likely to reveal one point of data (either the user understands or does not understand the concept in question), even a short multi-sentential piece is going to offer many points of data per utterance. Every grammatical construct successfully or unsuccessfully used, from determiner choice to word order, provides information about the user. These points can be correlated to provide a map of those constructs consistently used, those which are experiencing variation, and those which are absent; therefore, even during the initial tutorial session, we are provided with a fairly rich source of explicitly-derived data about this individual, compared to what we could obtain from questioning the student. Relying on a subjective categorization of language ability from a teacher would also be less accurate, as it is difficult to classify discrete levels of achievement in this domain, judgments are likely to vary between instructors, and categories would translate roughly at best to tags on the myriad individual language features.

Once the user placement and initial notation has occurred, implicit information can also be obtained; if a given feature is highlighted as within the ZPD for a student, this implies indirectly that features indicated as adjacent by lateral links to the other hierarchies are also in the ZPD, and that features above or below the ZPD are unknown or well-known respectively. In the absence of direct evidence to contradict these conclusions, the user model allows for this inferencing to produce reasonably certain conclusions. This is how we plan to exploit the "stereotypic reasoning" suggested by (Wahlster

___
[3]Unfortunately, the first session of error identification has to proceed without the assistance of a user model, but it will be aided by other data such as the expected co-occurrence of certain errors mentioned earlier.

and Kobsa, 1986), since the feature organization in SLALOM is based on a stereotypic acquisition order.

Note that we do not recommend explicitly marking the inferred knowledge in the model; following the lead of other explanation systems, implicit information in the user model can be derived at any time through inferencing, and thus should not be marked in the model so that it may be distinguished from explicit and confirmed information. Stereotypic information may not hold true for every individual, and we wish to distinguish between that information which we know from actual user performance and that which we infer from our profile of a typical learner.

When the system makes reference to this model during tutorial response generation, it must take note of whether it is drawing from an explicit or implicit source; the implicit information is less reliable, and our planner will again follow the conventions of similar systems and mark such inferences directly in its text plan for the purposes of recovery should they turn out false. It is hoped that the use of implicit user model information will be constrained to the early sessions with a given user only, since as argued above the ICICLE user model should be rich with explicit information and should be well-filled with direct information from user input in very few sessions. However, with revision of the representation over time, some tags may become less certain and the ability to infer additional information may be useful.

5.2 Updating the Model

In ICICLE, the responsibility for updating the model of a user lies with the error identification module, since that facet of the system processes all of the major parts of user performance. Each new analysis provides new (and potentially different) information that should be directed to the model. Because the user's knowledge is expected to change over time, so must the model. There is also the possibility that the user model is incorrect; even a rich model such as the one proposed for ICICLE may contain faulty data, so the system must be capable of revising earlier notations.

A model that can be overwritten over time gives rise to the question of whether new data should always champion over the old. The out-

line given thus far of what observations affect the model is fairly vague: features used "consistently correctly" are mastered concepts, those used "consistently incorrectly" are at the current level, and those which are rare or absent from the student's writing are above his or her level. The judgment of whether something is at or below a student's level may change when the amount of data increases as the system goes through more than one piece of the student's work, particularly if one or more of the pieces is too short to contain several instances of the feature. Therefore, it makes sense for the model to track certain figures (the number of times a feature is attempted, and the number of times it is executed without error) across more than one piece of writing and to make distinctions between figures collected within the most recent piece of writing and those collected across others in the past (since the user's proficiency will not change within a given piece, but there may be change across a selection of them). This will allow the system to examine as much data as possible, strengthening its ability to make these judgments. In this view, the user's writing is seen as a continuum of performance events over time from the first session to the most recent. But since the user's proficiency is also changing, the system should not always compute performance statistics which include events stretching back to the beginning of his or her use of the system, when the performance levels may have been different. Therefore, we recommend that the system maintain a "sliding window" of performance history across writing samples from which to update the user model at each new analysis. Ideally, this window would include enough data to be robust, and yet be small enough to capture only the "current" statistics. This latter requirement is particularly important for the system's self-evaluation and deciding whether recent explanatory attempts have succeeded. Determining what size such a window should be is a realm of future research. Related issues are whether or not it should adjust its size according to the circumstance, and what statistics of successful execution would be sufficient for judging a feature to be "consistent" in its use.

6 Conclusions

In this paper, we have discussed a proposed architecture for a model of user second language proficiency for use in a second language learning system. Our design is based on theories of second language acquisition research, particularly those involving interlanguage progression and concepts of orders of acquisition. We have also based the design on theories of cognitive skill acquisition by Vygotsky and others. The architecture we propose consists of hierarchies representing linear progressions of language feature acquisition, connected by lateral relationships of concurrent learning. Finally, we have shown that such a model, supported by preliminary empirical findings, can provide a rich and robust information base to a language assessment / correction application by modeling user proficiency at a high level of granularity and specificity.

References

N. Bailey, C. Madden, and S. D. Krashen. 1974. Is there a 'natural sequence' in adult second language learning? *Language Learning*, 24(2):235–243.

C. Baker and D. Cokely. 1980. *American Sign Language: A Teacher's Resource Text on Grammar and Culture*. TJ Publishers, Silver Spring, MD.

Giuseppe Carenini and Johanna D. Moore. 1993. Generating explanations in context. In Wayne D. Gray, William E. Hefley, and Dianne Murray, editors, *Proceedings of the International Workshop on Intelligent User Interfaces*, pages 175–182, Orlando, Florida, January 4-7. ACM Press.

Michelene T. H. Chi, Robert Glaser, and Ernest Rees. 1981. Expertise in problem solving. In *Advances in the Psychology of Human Intelligence*, chapter 1, pages 7–76. Lawrence Erlbaum, Hillsdale, NJ.

Vivian Cook. 1993. *Linguistics and Second Language Acquisition*. Macmillan Press Ltd, London.

Heidi C. Dulay and Marina K. Burt. 1974. Errors and strategies in child second language acquisition. *TESOL Quarterly*, 8(2):129–136, June.

Heidi C. Dulay and Marina K. Burt. 1975. Natural sequences in child second language acquisition. *Language Learning*, 24(1).

Rod Ellis. 1994. *The Study of Second Language Acquisition*. Oxford University Press, Oxford.

David Ingram. 1989. *First Language Acquisition: Method, Description, and Explanation*. Cambridge University Press, Cambridge and New York.

Stephen Krashen. 1981. *Second Language Acquisition and Second Language Learning*. Pergamon Press, Oxford.

Kathleen F. McCoy and Lisa N. Masterman (Michaud). 1997. A tutor for teaching English as a second language for deaf users of American Sign Language. In *Proceedings of Natural Language Processing for Communication Aids, an ACL/EACL97 Workshop*, pages 160–164, Madrid, Spain, July.

Kathleen F. McCoy, Christopher A. Pennington, and Linda Z. Suri. 1996. English error correction: A syntactic user model based on principled mal-rule scoring. In *Proceedings of the Fifth International Conference on User Modeling*, pages 59–66, Kailua-Kona, Hawaii, January 2-5. UM96, User Modeling, Inc.

Lisa N. Michaud and Kathleen F. McCoy. 1998. Planning text in a system for teaching English as a second language to deaf learners. In *Proceedings of Integrating Artificial Intelligence and Assistive Technology, an AAAI '98 Workshop*, Madison, Wisconsin, July.

Johanna D. Moore and Cecile L. Paris. 1992. Planning text for advisory dialogues: Capturing intentional and rhetorical information. *Computational Linguistics*, 19(4):651–695.

J. Michael O'Malley and Anna Uhl Chamot. 1990. *Learning Strategies in Second Language Acquisition*. Cambridge University Press, Cambridge.

Cecile Laurence Paris. 1987. *The Use of Explicit User Models in Text Generation*. Ph.D. thesis, Columbia University.

Martin H. Ringle and Bertram C. Bruce. 1981. Conversation failure. In W. G. Lehnert and M. H. Ringle, editors, *Knowledge Representation and Natural Language Processing*, chapter 7, pages 203–221. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Robert Rueda. 1990. Assisted performance in writing instruction with learning-disabled students. In Luis C. Moll, editor, *Vygotsky and Education: Instructional Implications and Applications of Sociohistorical Psychology*, pages 403–426. CUP, Cambridge.

David Schneider and Kathleen F. McCoy. 1998. Recognizing syntactic errors in the writing of second language learners. In *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and the Seventeenth International Conference on Computational Linguistics*, volume 2, pages 1198–1204, Universite de Montreal, Montreal, Quebec, Canada, August 10-14. COLING-ACL, Morgan Kaufmann Publishers.

Bonnie D. Schwartz and Rex A. Sprouse. 1996. L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12(1):40–72.

Bonnie D. Schwartz. 1998. On two hypotheses of "transfer" in L2A: Minimal Trees and Absolute L1 Influence. In S. Flynn, G. Martohardjono, and W. O'Neil, editors, *The Generative Study of Second Language Acquisition*, chapter 3, pages 35–59. Lawrence Erlbaum, Mahwah, NJ.

L. Selinker. 1972. Interlanguage. *International Review of Applied Linguistics*, 10:209–231.

Linda Z. Suri and Kathleen F. McCoy. 1993. A methodology for developing an error taxonomy for a computer assisted language learning tool for second language learners. Technical Report TR-93-16, Dept. of CIS, University of Delaware.

Lev Semenovich Vygotsky. 1986. *Thought and Language*. MIT Press, Cambridge, MA.

Wolfgang Wahlster and Alfred Kobsa. 1986. Dialog-based user models. In Giacomo Ferrari, editor, *Proceedings of the IEEE, Special Issue on Natural Language Processing*, February.

Beverly P. Woolf. 1984. Context dependent planning in a machine tutor. Technical Report COINS Technical Report 84-21, Dept. of Computer and Information Science, University of Massachusetts at Amherst, May.