Supporting Intelligent Tutoring in CALL By Modeling the User's Grammar

Lisa N. Michaud and Kathleen F. McCoy

Computer and Information Sciences Dept.
University of Delaware, Newark, DE 19716
{michaud,mccoy}@cis.udel.edu
http://www.eecis.udel.edu/research/icicle

Abstract

This paper presents a model for representing the proficiency of users in a CALL system by recording their performance on specific grammatical features. The model will be used both to select accurate interpretations of user-written sentences and to focus system-delivered instruction on topics at the frontier of the learner's competence.

Introduction

Modeling characteristics of the user in an intelligent tutoring system is an essential undertaking if the system is to adapt itself to the needs of the individual learner. It is, after all, the adaptability of a tutoring system which distinguishes it from an instructional text, whose only capability is to provide information to the learner in a predetermined manner regardless of the learner's existing knowledge or strengths. A well-designed tutoring system plays two roles: it is a diagnostician, discovering the nature and extent of the student's knowledge, and a strategist, planning a response (such as the communication of information) using its findings about the learner (Glaser, Lesgold, & Lajoie 1987; Spada 1993). A model of the user typically serves as a repository for the information passing between these two processes, representing what has been discovered about the learner and making that data available to drive the decisions of the system when planning tutorial actions.

It has been argued (Sparck Jones 1991; Cawsey 1993) that creating and maintaining a detailed user model in a system involving natural language interaction is a very difficult task. Sparck Jones in particular argues that evidence for user modeling in such a system is likely to be "poor in both quantity and quality," and that "fancy modeling chasing the real person is unnecessary." We argue that this pragmatic conservatism is not universally appropriate. In our system, a Computer Assisted Language Learning

(CALL) system which instructs on English as a second language through the paradigm of a writing tutor, we have developed an architecture for a detailed model of the user's language competence to be used both when interpreting the linguistic input to the system and when selecting the topics for the instructional material. We hold that this model contains a high level of detail and yet is robustly supported by the input available to the system. In this paper, we present an overview of our system and address part of its interaction with this component of the user model, which is the focus of current research.

The ICICLE System

The name ICICLE stands for "Interactive Computer Identification and Correction of Language Errors" and is the name of an intelligent tutoring system under development (Michaud & McCoy 1998; Schneider & McCoy 1998; Michaud & McCoy 1999). Its primary goal is to employ natural language processing and generation to tutor deaf users of American Sign Language on their written English grammar. Of paramount importance to this goal is the correct analysis of the source and nature of user-generated language errors and the production of tutorial feedback to student performance which is both correct and individualized, taking into account the language knowledge, proficiency, and learning style of the student.

ICICLE's interaction with its user takes place primarily through a cycle of user input and system response. The cycle begins when a user submits a piece of writing to be reviewed by the system. The system then determines the grammatical errors in the writing, and constructs a response in the form of written tutorial feedback. This feedback is aimed toward making the student aware of the nature of the errors found and toward giving him or her the information needed to correct them. When the student makes those corrections and/or other revisions to the piece, it is re-submitted for analysis and the cycle begins again. As ICICLE is intended to be used by an individual over time and across many pieces of writing,

the cycle will be repeated many times.

An Effective ESL Tutor

The current status of the ICICLE system is a functioning text parser using an English grammar augmented by "mal-rules" which capture typical errors made by our learner population (Schneider & McCoy 1998). It has the ability to recognize and label grammatical errors, delivering "canned" one- or two-sentence explanations of each error on request.

In operation, when the system finds more than one possible analysis of a user's sentence,[1] it currently chooses the first error-free analysis of the list, if any, or the first of all of the parses if there is no grammatical possibility. In order to make more principled choices in this selection, and to enable a more complex tutorial component which will provide original text generation communicating instruction tailored to the individual, it requires the addition of a user model whose contents illustrate multiple characteristics of the student user. The multi-component model we have designed incorporates elements to track the history of the user's interaction with the system in addition to models of the user's grammar proficiency and conscious domain knowledge. For the purposes of this paper, we will address the grammar proficiency component of the model, which is directly supportive of the error identification process and the initial stage of tutorial session planning.

A Model of the User's Grammar

This component is called SLALOM[2] and it involves a representation of the user's ability to correctly use each of the grammatical "features" of English. These features include aspects of grammar such as pluralizing a noun with +S, or making appropriate use of the past tense. The information stored about each of these features represents the observations made by the system based on the performance it has observed over the submission of multiple pieces of writing by a given user. If the user typically uses a given feature correctly, its corresponding element in the model will be marked "acquired." Conversely, consistent violation of a grammar rule will cause it to be marked "unacquired." We also wish to represent a third realm of proficiency, based on Vygotsky's observations about the acquisition of cognitive skills. He used the term *Zone of Proximal Development* (ZPD) to capture that subset of the skill which the learner is about to master (Vygotsky 1986). Krashen's observation that at each step of language learning there

is some set of grammar rules which the learner is "due to acquire" (Krashen 1982) effectively applies this theory to our domain.

Ellis (Ellis 1994) helps us determine which features are in the ZPD by characterizing the nature of grammatical structures on the verge of acquisition; he observes that those which are about to be acquired tend to exhibit variation in use, some of which is grammatically inappropriate in its syntactic context, before usage settles down. Observation of inconsistent behavior should therefore clearly flag for ICICLE those features which should be marked "ZPD" for this learner.[3]

Accurately Diagnosing Student Errors

When processing a student's writing, one of the ICICLE system's primary tasks is to obtain accurate analyses of ungrammatical text. As addressed earlier, in some cases there exist multiple structural possibilities. The differentiation between these possibilities may depend upon the proficiency of the learner; advanced language learners typically make different errors than novice ones. Moreover, an individual will make errors on different aspects of English as his or her proficiency develops and certain concepts are mastered. Therefore, we intend to enable the system to choose between the structural analyses using SLALOM's information about the user's language proficiency.

An established set of SLALOM tags should enable ICICLE's error identification process to proceed on the premise that future user performance can be predicted based on the patterns of the past. If the tags have been assigned in the model based on the performance of the user to date, then if a feature has been marked "acquired," the user tends to execute it correctly, whereas a feature marked "unacquired" indicates that it is usually broken; the system can therefore generally prefer parses which use rules representing well-formed constituents associated with "acquired" features, mal-rules from the "unacquired" area, and either correct rules or mal-rules for those features marked "ZPD." Figure 1 shows how different tags on the same item in the model might "highlight," or make preferable, certain rules in the parsing grammar.[4] If the system is then attempting to analyze the sentence, "My brother like baseball," and the model indicates that the user's mastery of subject/verb agreement is well-established but his mastery of plural nouns is not, it can prefer a parse con-

---

[1]There is sometimes more than one possible parse, potentially resulting in different errors being assigned.

[2]The meaning of the name is discussed later in this paper.

[3]Some features may not appear in the student's writing for some time; this will be addressed later.

[4]Represented via description; the actual rules of the grammar are LISP constructions and are not easily readable.
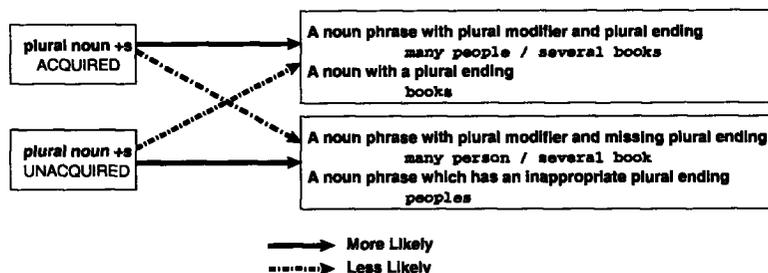
Figure 1: Model tags highlighting rules in the grammar.

taining a mal-rule which marks "brother" as a plural noun which is missing an "+S" ending over an interpretation in which agreement is missing from the verb.

Focusing on the Frontier of Learning

Once the text has been analyzed, ICICLE must generate a tutorial session to address the errors it has found. It will begin by determining which of the errors will be the subjects of tutorial explanations. This decision is important if the instruction is to be effective, for the learner has a narrow zone of topics which are appropriate for instruction. Some of the perceived ungrammaticalities in the text are not actually representative of user competence. The distinction between second language "errors," which reflect grammatical competence, and "mistakes," which are merely slip-ups, has been addressed by researchers such as (Corder 1967) and is of high relevance to a tutoring approach which endeavors to avoid unnecessary instruction. If the ungrammaticality is simply a mistake, ICICLE should mark it but exclude it from being addressed by tutorial actions, since the learner already possesses that knowledge.

We also want to avoid generating instruction which would go over the student's head. This is partly due to common sense and partly due to the concern that second language instruction cannot result in fully assimilated knowledge if not constrained by "learnability" concerns, under which a learner cannot acquire the knowledge if he or she is not developmentally prepared to do so (Ellis 1993). It is our intent therefore to focus instruction on the ZPD, that "narrow shifting zone dividing the already-learned skills from the not-yet-learned ones" (Linton, Bell, & Bloom 1996), or the frontier of the learning process, since instruction outside of this area may not result in learning and is wasteful of time and effort. ICICLE will select those errors which involve features from this learner's ZPD and use them as the topics of its tutorial feedback.

Partial Evidence in SLALOM

We have shown how our model of tagged grammatical features representing past performance may facilitate decisions between analyses of user-written sentences and enable ICICLE to focus its tutoring efforts. One problem with this approach is the necessity of making judgments on user competence from incomplete information; ICICLE will not always have empirical data covering all features in SLALOM. An unacquired feature may have been absent due to avoidance, or an acquired feature absent due to lack of opportunity. We therefore must establish a method by which the system can infer a fuller description of user proficiency than is directly displayed in his or her past use of language forms.

SLALOM (*Steps of Language Acquisition in a Layered Organization Model*) possesses a structure designed to help ICICLE fill in the gaps. A very simplified representation of SLALOM can be found in Figure 2, where each box represents a grammar feature. *Steps of Language Acquisition* refers to our intention to capture the order of acquisition of second language features. There is empirical support for stereotypical sequences of language acquisition, and we wish to represent this by ordering features in our model according to these sequences. In particular, SLALOM groups them into "hierarchies" of related features (such as morphology markings, NP constructions, and relative clause formation), each of which
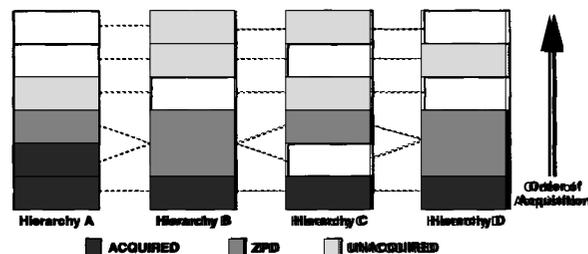

Figure 2: SLALOM: Steps of Language Acquisition in a Layered Organization Model.
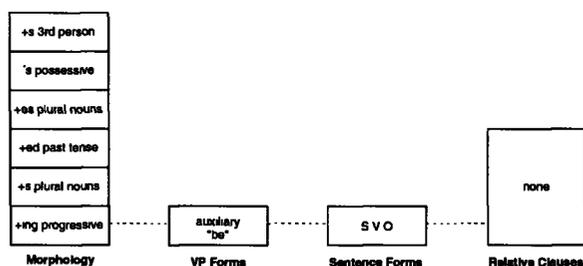
Figure 3: Examples of a SLALOM hierarchy and layer.

has an order represented in the figure by a vertical relationship; "easier" features which are typically acquired earlier sit below those acquired later.

The *Layered Organization Model* part of SLALOM's design is shown in the figure by the dashed lines connecting elements across the hierarchies. These connections serve both to coordinate the acquisition steps across the hierarchies and to indicate "layers" of concurrent acquisition; elements connected at the same layer are acquired at about the same time. Intuitively, at some moment in the learner's acquisition process, one layer is the current ZPD; these items are being acquired at the present time. Typically, those items below that layer have already been acquired, while those above have not been acquired.[5]

Figure 3 demonstrates a possible SLALOM hierarchy and layer. The morphology hierarchy is based on the results of (Dulay & Burt 1975), who found that learners of English as a second language typically learn +*ing* progressives before +*s* plurals before +*ed* past tense, etc. The layer indicates that +*ing* is mastered about the same time as the learner acquires auxiliary "be" in VPs and S V O sentences, while relative clauses have not appeared yet.[6] Statistical analysis on a corpus of 101 samples of writing by deaf students yielded preliminary results showing different sets of errors committed by different levels of ability, while certain errors co-occur at the same level in significant degree. We intend to supplement this data with further analysis, existing order of acquisition work, and a longitudinal study of learners from the target population in order to establish prototypical acquisition relationships.

---

[5]Note that what is considered a "layer" may be much larger than just one item per hierarchy. The important aspect of the definition is that each layer represents a grouping of language rules acquired at about the same time.

[6]This layer is for example purposes only and does not reflect any empirical findings, but current efforts are addressing this.

Although these relationships will be based on a general learner profile and not on the individual, they can serve to supplement the solid data we have on a specific learner. If an item in SLALOM has not yet received a tag, but it is below those items marked "ZPD" in SLALOM (or perhaps even below those that are "acquired"), it should be considered acquired. Likewise, one above the ZPD or above "unacquired" structures should be considered unacquired. Those at the same layer as the ZPD should also be part of it.

Once a user has begun to attempt a given construction, whether successfully or unsuccessfully, his or her performance will determine the marking on that construction in SLALOM and the model's organization will be irrelevant. The system therefore only has to rely on stereotypical data in novel situations; if a learner is acquiring features out of order due to instructional emphasis in the classroom, then his or her markings will reflect this and the system's decisions will be based on the individual, not the population of second language learners as a whole.

A Dynamic Model

SLALOM's tags will be initialized following the first performance analysis of a new user's writing. Those features he or she has used consistently correctly will receive "acquired" tags, those used incorrectly "unacquired" tags, and those in variation "ZPD" tags. With each analysis of a new piece of writing from the student, these observations will be augmented with new and potentially different data, as features originally tagged as part of the ZPD exhibit correct usage and features originally tagged "unacquired" begin to show signs of variation and move into the ZPD. New data will result in the SLALOM tags being revised to reflect the user's developing knowledge. Because SLALOM represents an expected order of acquisition, the likely path of the ZPD would be to move "up" in the stacks.

We are aware of the difficulty of performing accurate parses in the initial evaluation of a user without the support of SLALOM's tags. We intend to investigate a two-pass approach, where a first pass through the piece evaluates such crude measures of writing competence as mean number of words per utterance or complexity of clause structure. Using observations from the first pass to give it a general idea of a global user competence level, ICICLE would be able to make initial decisions that are not entirely arbitrary.

Regardless of how SLALOM receives its initial markings, however, it is clear that its accuracy will improve greatly over time. Because the system is intended to be used by an individual over many pieces of writing, it will have access to a continually growing

corpus of user-produced utterances. ICICLE is not subject to the same limitations as the dialogue systems on which Sparck Jones based her observations in (Sparck Jones 1991) for two reasons: first, the number of user utterances it has access to is much larger because they are not artifacts of natural interaction but fed to the system in large batches; and second, the user knowledge that is measured by SLALOM is not that which is *communicated* by these utterances (semantic content), but that which is *exhibited* by the utterances (syntactic content). This both increases the data extracted from each sentence and removes a lot of ambiguity, making it a far more accessible task for a machine to judge the extent of user knowledge.

Summary

We have focused in this paper on the need for a representation of a learner's grammatical proficiency in the ICICLE system, and have briefly addressed how the design of this model (under development) will interface with the existing system and the future tutorial component. Because our system reviews multiple pieces of writing from a given user over time, it is feasible to argue that performance evaluation derived from this review process could feed the language proficiency model with robust, changing data about what aspects of the grammar the user has mastered and which are still causing him or her the most difficulty. Ongoing research will address the exact architecture of the grammar model design and related implementation issues. Our goal is to illustrate how a model thus constructed can aid a CALL system in obtaining accurate interpretations of student performance and help to support a complex and effective tutoring planner.

Acknowledgments

References

Cawsey, A. 1993. *Explanation and Interaction: The Computer Generation of Explanatory Dialogues.* Cambridge, MA: MIT Press.

Corder, S. P. 1967. The significance of learners' errors. *International Review of Applied Linguistics* 5(4):161–170.

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

Ellis, R. 1993. The structural syllabus and second language acquisition. *TESOL Quarterly* 27(1):91–113.

Ellis, R. 1994. *The Study of Second Language Acquisition.* New York: Oxford University Press.

Glaser, R.; Lesgold, A.; and Lajoie, S. 1987. Toward a cognitive theory for the measurement of achievement. In Ronning, R. R.; Glover, J. A.; Conoley, J. C.; and Witt, J. C., eds., *The Influence of Cognitive Psychology on Testing,* volume 3 of *Buros-Nebraska Symposium on Measurement and Testing.* New Jersey: Lawrence Erlbaum Associates. chapter 3, 41–85.

Krashen, S. D. 1982. *Principles and Practice in Second Language Acquisition.* New York: Pergamon Press.

Linton, F.; Bell, B.; and Bloom, C. 1996. The student model of the LEAP intelligent tutoring system. In *Proceedings of the Fifth International Conference on User Modeling,* 83–90. Kailua-Kona, Hawaii: UM96.

Michaud, L. N., and McCoy, K. F. 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.*

Michaud, L. N., and McCoy, K. F. 1999. Modeling user language proficiency in a writing tutor for deaf learners of english. In Olsen, M. B., ed., *Proceedings of Computer-Mediated Language Assessment and Evaluation in Natural Language Processing, an ACL-IALL Symposium,* 47–54. College Park, Maryland: Association for Computational Linguistics.

Schneider, D., and McCoy, K. F. 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, 1198–1204. Universite de Montreal, Montreal, Quebec, Canada: COLING-ACL.

Spada, H. 1993. How the role of cognitive modeling for computerized instruction is changing. In Brna, P.; Ohlsson, S.; and Pain, H., eds., *Proceedings of AI-ED'93, World Conference on Artificial Intelligence in Education,* 21–25. Edinburgh, Scotland: Association forthe Advancement of Computer in Education (AACE). Invited talk.

Sparck Jones, K. 1991. Tailoring output to the user: What does user modelling in generation mean? In Paris, C. L.; Swartout, W. R.; and Mann, W. C., eds., *Natural Language Generation in Artificial Intelligence and Computational Linguistics.* Boston: Kluwer Academic Publishers. chapter 8, 201–225.

Vygotsky, L. S. 1986. *Thought and Language.* Cambridge, Mass.: The MIT Press.