

Capturing the Evolution of Grammatical Knowledge in a CALL System for Deaf Learners of English

Lisa N. Michaud, *Department of Mathematics and Computer Science, Wheaton College, Norton, MA 02766, USA*
lmichaud@wheatoncollege.edu

Kathleen F. McCoy, *Department of Computer and Information Sciences, University of Delaware, Newark, DE 19716, USA*
mccoy@cis.udel.edu
<http://www.eecis.udel.edu/research/icicle>

Abstract. The ICICLE project is a Computer-Assisted Language Learning (CALL) environment geared toward teaching English as a second language. This paper reports on an initial prototype application of the system for deaf learners of written English. A primary focus of the ICICLE effort has been devoted to enabling the system to adapt to a learning user over the evolution of language proficiency. In this paper, we overview and motivate the design of our novel user modeling component that integrates Selinker's Interlanguage theory and other research in Second Language Acquisition in order to accurately represent a learner's internal grammar as it changes over time. The objectives of this effort are two-fold: to accurately diagnose and respond to learner errors, and to focus tutorial feedback on those errors which are most relevant to the learner's acquisition process.

Keywords. Intelligent Tutoring Systems, CALL, natural language processing, student modeling

INTRODUCTION

We are currently developing the **ICICLE** system, a Computer-Assisted Language Learning (CALL) system which instructs on English as a second language through the paradigm of a writing tutor (Michaud, 2002; Michaud & McCoy, 2004; Michaud *et al.*, 2005). The name ICICLE represents "Interactive Computer Identification and Correction of Language Errors." The system's long term goal is to employ natural language processing and generation to tutor learners of English as a second language on grammatical components of their written texts. Our work to date toward this goal has focused on the correct analysis of the source and nature of user-generated language errors so that the system can generate tutorial feedback to student performance which is both correct and individualized.

ICICLE's interaction with the user is intended to take place over long periods of time as various pieces of writing are submitted to the system for analysis. The interaction concerning a single piece

of writing is accomplished through a cycle of user input and system response. This cycle begins when a user submits the piece of writing to review by the system; the system performs an analysis on this writing, determines its grammatical errors, and constructs a response in the form of tutorial feedback. This feedback is aimed toward making the student aware of the nature of the errors found in the writing 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 resubmitted for analysis and the cycle begins again.

Although ICICLE is a general framework and can be adapted to any population of English learners, our current implementation has been designed to be used with deaf students. Below we overview part of our motivation for selecting this particular user population.

Education Issues for the Deaf

Literacy figures for deaf students are poor to the extent of being shocking (c.f. Paul, 1998; Quigley & King, 1982, for discussions of empirical studies), despite the fact that intelligence is distributed normally in this population (Swisher, 1989). With this problem impacting every aspect of a deaf student's education and future (Loera & Meichenbaum, 1993; Moores, 1987), the search for an approach to improving the situation demands a second look at the unique needs of this learner group and at how a natural language system may be designed to meet them.

American Sign Language (ASL) is a communication form used by many deaf individuals; Charrow & Wilbur (1975) reported that at the time of their writing, ASL was the third most widely-used non-English language in the United States after Spanish and Italian. They also reported that most prelingually deaf adolescents and adults in the United States have ASL as their native language. We have therefore chosen to focus on native or near-native users of ASL as our target learner group for the ICICLE system. While having a strong native language base is of great benefit in the acquisition of a second language, and observations suggest that those deaf individuals who have had the benefit of early and natural acquisition of ASL display increased aptitude for the acquisition of English (c.f. Charrow & Fletcher, 1974; Charrow & Wilbur, 1975; Swisher, 1989), the fact that they are indeed two distinct languages and the existence of broad differences between ASL and English together pose challenges for the learner attempting to transfer general language knowledge from one to the other. For instance, ASL is a visual-gestural language whose grammar is distinct and independent of the grammar of English or any other spoken language (Baker & Cokely, 1980). The sign-order rules of ASL are not the same as the word-order rules of English, and ASL syntax includes systematic modulations to signs as well as non-manual behavior (e.g., posture and facial expression) that achieve a simultaneous mode of communication not possible with the completely sequential nature of written English (Baker & Cokely, 1980).

We discussed these differences plus the differences in cognitive processing techniques between spoken and manual languages in (Michaud & McCoy, 1998) and (Michaud *et al.*, 2000). For ASL natives, English is a distinctly different and challenging language, motivating the need to view the process of acquiring fluency in written English as second language acquisition and to incorporate that view in a strategy for facilitating the learning process. By espousing this perspective, we are consistent with the "Bi-Bi" philosophy in deaf education, where deaf children are seen as being *bilingual* and *bicultural*. The primary mode of communication in a Bi-Bi classroom is ASL, and English is taught as a second

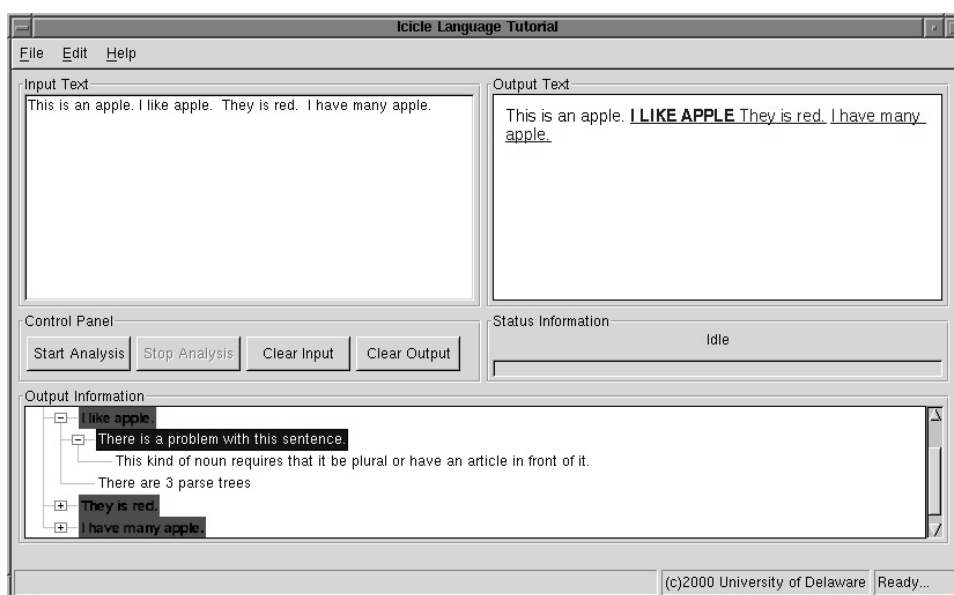


Fig.1. A screenshot of the current ICICLE implementation.

language (Barnum, 1984; Drasgow, 1993; Erting, 1978; Gutiérrez, 1994; Johnson *et al.*, 1989; Quigley & Paul, 1984; Stevens, 1980; Strong, 1988; Swisher, 1989). This philosophy encourages the students to build on their strong ASL language foundation as they acquire written English.

ICICLE therefore attempts to address the many difficulties facing the deaf user of ASL by providing an environment in which he or she can practice the usage of written English and get grammatical feedback and instruction without the “loss of face” associated with a human tutor. This target user group, however, is very heterogeneous (Stewart, 2001), spanning individuals who can produce utterances near to those of a native English speaker and individuals who struggle with basic English structures. In order to deal appropriately with such a population, it is clear that ICICLE needs to incorporate adaptivity to the level of its user.

Current Implementation and Architecture

The current prototype implementation of ICICLE is a graphical application which provides the learner with several windows through which to work with the text. The learner may load the text of an essay into the main window shown in the upper left of the screen shot in Figure 1. The learner may then ask the system to analyze the text, and, once analyzed, the text will be redisplayed in the window to the right with the errors identified via underlining. In a future implementation, a tutorial response employing natural language generation will explain the errors to the user; in this prototype, the bottom window of the application reproduces all sentences with errors and the user has the option of querying the nature of those errors, accessing canned one-sentence explanations. (The prototype also currently allows the user to access a graphical depiction of the parse trees for each sentence, something intended for the ICICLE

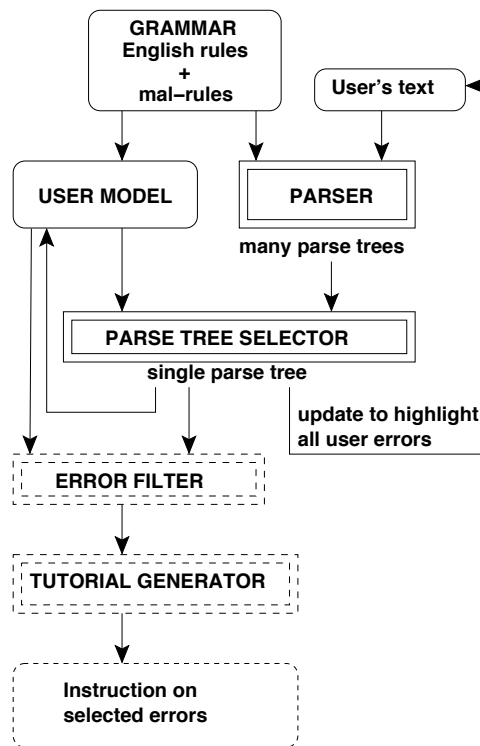


Fig.2. The ICICLE system architecture.

developers only that will not be openly available if the system is deployed among the target users.) The text may then be edited and reanalyzed as needed.

In order to identify errors, ICICLE uses a text parser to syntactically interpret samples of user-written text. The analysis process relies on lexical information from the COMLEX Syntax 2.2 lexicon (Grishman *et al.*, 1994), the Kimmo morphological processor (Karttunen, 1983; Antworth, 1990), and Allen's TRAINS parser version 4.0 (related to the parser in (Allen, 1995)). We have provided the parser with an augmented CFG grammar which has been developed specifically for ICICLE. This grammar currently consists of 321 rules representing standard English constructions, augmented by mal-rules (Sleeman, 1982), or bug rules, which represent commonly-committed grammatical errors, derived from an error taxonomy compiled out of actual writing samples from deaf college students (Suri & McCoy, 1993)¹.

Figure 2 depicts how the components of ICICLE work in concert to connect the work of this parser to the rest of the system. ICICLE analyzes each sentence of the learner's writing in turn using the parser. The output produced by this process is a (sometimes large) set of possible parse trees, each representing

¹The mal-rules in the grammar may be dependent on the first language of the learner to some extent. We anticipate that in adapting a future implementation of ICICLE to another population of learners, we may retain some of the existing mal-rules, while new mal-rules may need to be added, according to the typical errors committed by the new target population.

a different syntactic analysis of the sentence. Some of the potential parses for a given sentence may contain errors, and others may not; in addition, different parses may place the 'blame' for grammatical errors on different constituents. Selecting a single parse that best represents what grammatical structure the learner most likely intended to use has been a large focus of this work and is discussed below. Since determining the nature and cause of student errors is an integral step to deciding how to approach student instruction (Matz, 1982), the parser must be able to make principled decisions between these options.

The single parse that is most likely given the learner's current mastery of the language is the one selected by the Parse Tree Selector. If this parse used mal-rules, then the sentence will be highlighted via underlining in the output display. In the future, components shown with dashed lines will be fully implemented, so that an Error Filter will cull from all of the errors committed those which are most relevant to the learner's current acquisition status (as indicated by the User Model), and those errors will be passed to a Tutorial Generator so that natural language explanations of the errors may be provided. As mentioned earlier, the explanations currently provided are constructed of canned text, and no filtering has yet been applied to the production of explanations.

It has been said that a well-designed tutoring system actively undertakes two tasks: that of the diagnostician, discovering the nature and extent of the student's knowledge, and that of the strategist, planning a response (such as the communication of information) using its findings about the learner (Glaser *et al.*, 1987; Spada, 1993). A user model 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.

In terms of this description of an intelligent tutoring system, the receipt of a new piece of writing and its subsequent analysis is where ICICLE plays the role of the diagnostician; the planning of a tutorial response is where the role of the strategist is played out. Since the system is intended to be used by an individual over time and across many pieces of writing, these roles and the cycle they represent will be performed many times with a user; therefore, it is clearly beneficial for ICICLE to maintain a model of its users, serving as the repository for the information gathered by, and referenced for, each of those two tasks. With the aid of such a model, ICICLE adapts itself to the changing needs of a student across the learning journey.

Below, each step of ICICLE's cyclic interaction with its user is outlined in terms of what demands it places on a model of its user in order to function in an adaptive way.

Parse Disambiguation

As stated above, in order to obtain a correct analysis of the source and nature of student errors, the parse tree selection module of the ICICLE system needs to choose between multiple parses or interpretations of each utterance. This selection may significantly affect the feedback the system gives to the user because the different parses may place the blame for the ungrammaticality in different places. For example, the parser may encounter the following sentence in the writing of a student:

- (1) * She is teach piano on Tuesdays.

Different parses of this sentence may involve different error-capturing mal-rules exhibiting fundamentally different errors. We claim that having a model of the student's current mastery of English will

help disambiguate between these choices. For instance, the system must determine whether the student is perhaps a beginning learner who is over-applying the auxiliary *BE* and had intended to generate the simple present instead:

(2) [*Interpretation*] She teaches piano on Tuesdays.

Another possibility is that the student has mastered the morphology of the simple present but has trouble forming the morphology of the progressive tense:

(3) [*Interpretation*] She is teaching piano on Tuesdays.

A third possibility is that the student has actually mastered both present and progressive tenses but is struggling with forming passive voice:

(4) [*Interpretation*] She is taught piano on Tuesdays.

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 grammatical proficiency which indicates his or her mastery of such language concepts as the morphology of the present, progressive, and passive forms of verbs. This knowledge aids in choosing between structurally-differentiated parses by providing information on which grammatical constructs the user can be expected to use correctly or incorrectly (McCoy *et al.*, 1996)².

Selective Tutoring

In a future version of the ICICLE system, after identifying the errors committed by a user, the system will select a subset of these errors to send to the tutorial response component for the generation of instructive text. This will be done by the Error Filter, which will rely heavily on the user model component. While all errors that ICICLE identifies will be marked in the output window, we intend ICICLE to give instruction only on those language components which are at the user's *current level* of acquisition; errors on those above this level are likely to be beyond the user's understanding, while errors on those which are well-established are likely to be simple mistakes which do not require instruction. Generating a tutorial response to errors in either of these ranges is likely to cause frustration in the learner (Linton *et al.*, 1996). At the same time, many researchers agree (c.f. Corder, 1967; Rueda, 1990; Vygotsky, 1986) that the efficacy of second language (L2) instruction is greatly increased when the instruction is adapted to the needs of the learner in his or her current state of acquisition. Learnability theory constrains the ability of L2 learners to acquiring only those features they are ready to learn (Bialystok, 1978; Ellis, 1993; Higgins, 1995), and specifically in the case of our target learner audience, Kelly (1987) argues that pointing out every error committed by the deaf writer has the potential to be counter-productively overwhelming.

These are all very strong arguments for the system to target its instruction selectively. Therefore, when ICICLE decides what user errors should provoke instruction, it should narrow this choice to that

²This is not to say that the user will not make mistakes in already-mastered material. What we wish to produce is the most likely parse given the current mastery of the language.

“narrow shifting zone dividing the already-learned skills from the not-yet-learned ones” (Linton *et al.*, 1996, p. 83). Focusing instruction on this range has been the goal of other instruction systems such as Meno-tutor (Woolf, 1984; Woolf & McDonald, 1984), MULEDS (Opwis, 1993; Ploetzner *et al.*, 1990), and LEAP (Linton *et al.*, 1996).

The need to be selective when deciding what should be tutored upon places an additional demand on the user model: not only must it show the user’s command of each grammatical structure, but it also must indicate which structures are likely to be learned next. With such knowledge, the error analysis component may trim away those errors outside this indicated realm of accessible and productive learning.

A VIEW OF SECOND LANGUAGE ACQUISITION

So far, we have established that a user model that indicates (1) a student’s current level of grammatical competence and (2) what grammatical structures the user is currently attempting to acquire would greatly benefit a CALL system. We turn now to some linguistic theories of the Second Language Acquisition process as a basis for our user model’s design.

Interlanguage

In previous work on the ICICLE user model, we established its essential nature as a representation of the user’s location along the path toward acquiring written English as a second language (Michaud & McCoy, 1999, 2000). Corder (1973) stressed the need for such a representation to go beyond the question, “Does the learner know structure X ?” to model the answer to: “What language rules is the learner using?” To design a representation to contain such information, we turned to the Interlanguage theory of Second Language Acquisition (Selinker, 1971, 1972). In this theory, a learner produces utterances in the L2 from an internalized grammar which represents his or her hypothesis of the L2. From an initial hypothesis I_1 , the interlanguage is revised systematically over time as the learner acquires the L2, progressively moving toward a model which results in more native-like linguistic performance (Corder, 1974). The target of these revisions is some I_n , an approximation of the L2³.

A Language Hypothesis

If the learner’s interlanguage were captured at a given moment in this progression, it would be a “complete” grammar of the L2 as the learner understands it at this time. Corder (1974) referred to this individual language hypothesis as a kind of “transitional” dialect which he labeled an “idiolect.” Somewhere between the learner’s earliest hypothesis I_1 and the target goal I_n , the stage I_i is therefore a distinct interlanguage in and of itself (Schwartz, 1998; Schwartz & Sprouse, 1996), and is the grammar from which the learner is currently generating L2 utterances.

³If the learning were perfect, the resulting grammar would be that of the L2. However, many researchers believe that the L2 acquisition is almost always imperfect and that rule fossilization often occurs prior to actually achieving the L2, therefore resulting in an imperfect approximation.

The core of ICICLE's user model, a representation of current user language competence, strives to capture this I_i — a moment of the user's internalized grammar. The contents of this model reveal to the system the status of the learner's acquisition of the grammatical structures recognized by the system.

Interlanguage Transitions

Although we may speak of I_i , a given stage of a learner's interlanguage, we must also acknowledge that the interlanguage is not a static entity. As a hypothesis, it is under constant revision as the learner systematically examines portions of the hypothesis and updates them to reflect the L2 more closely. Over time, more of the interlanguage correctly models the target language, and less reflects incorrect assumptions.

Many researchers refer to this transitional process as Hypothesis Testing, where the learner is actively engaged in the systematic comparison of a portion of his or her interlanguage hypothesis against the L2, inducing new rules and testing their validity (c.f. Corder, 1974)⁴. It is understood that the learner perceives a difference between his or her productions (generated by his or her internal grammar, the interlanguage) and the input which he or she receives, leading to a revision or restructuring of the interlanguage grammar in favor of the target-like form (Brown & Hanlon, 1970; Carroll, 1995; Selinker, 1971, 1972, 1992). Ellis (1994) characterizes each "step" of the interlanguage grammar as sharing "rules" with the previous step, but differing in that some rules have been added or revised. This revision affects the portion of the interlanguage hypothesis that is currently the focus of the learner's Hypothesis Testing.

The Frontier of Acquisition

As interlanguage transitions occur between states I_i and I_{i+1} , this portion of the hypothesis which is the focus of the learner's attention is of interest to us because it represents those language structures the learner is currently in the process of acquiring, the very set of structures we discussed above as the ideal focus of effective tutorial instruction because they are neither already mastered nor beyond the user's reach.

We have adapted Lev Vygotsky's concept of a subset of material being mastered that is currently within the learner's grasp to acquire, the Zone of Proximal Development (ZPD) (Vygotsky, 1986). The general idea has been already applied to second language acquisition by researchers such as Washburn (1994) and Krashen (1982), who stated that when the learner is at some level i in acquiring the L2 grammar, there is some part of the grammar at level $i + 1$ that the learner is "due to acquire". Applying the ZPD concept directly to the interlanguage domain, therefore, the ZPD corresponds to the portion of the interlanguage that is the center of Hypothesis Testing and is in the process of making a transition to the target (L2) grammar.

The identification of the ZPD for a given second language learner would be an ideal indication of the next language structures he or she will acquire — the "frontier" of the acquisition process — and

⁴Note that the selection of the portion of grammar to come under Hypothesis Testing is systematic and not a random choice. The uniformity of the sequence of acquisition across different learners is discussed in more depth later.

consequently, those structures on which instruction would be most beneficial because they are neither well-established nor beyond his or her ability to learn at this time.

Researchers in the acquisition of learner knowledge in general (c.f. Ragnemalm, 1996) and language acquisition in particular (c.f. Brown & Hanlon, 1970; Corder, 1967; Ellis, 1994; Krashen, 1983) have noted the “immature” or “transitional” performance which appears for a time before acquisition. From this it is possible to conclude that the changing errors observed in second language learners are indicative of the frontier of their acquisition process. While a single error may be relatively meaningless, the consistent occurrence of specific errors gives us information from which we can describe “a picture of the linguistic development of a learner” (Corder, 1974, p. 125), or the learner’s current interlanguage state (I_i).

Toward Modeling the Interlanguage

We stated above that by modeling the user’s language proficiency, we would empower the ICICLE system to discriminate between competing syntactic parses of user utterances on the basis of what L2 performance can be expected from this user. In order to describe how this would proceed, we must first establish the relationship between our user model representing a learner’s interlanguage state I_i and the parses between which ICICLE must choose. These parses are made of syntactic constituents formed by rules in our parsing grammar. As discussed above, this grammar consists of rules reflecting both standard and error-containing English syntax. In effect, each “language structure” in the user’s interlanguage is realized by some group of specific syntactic rules representing different possible realizations, both correct and incorrect.

As the learner forms and tests an interlanguage hypothesis, he or she focuses on certain structures and tests the hypothesis concerning these structures. If learning is completely successful, as acquisition progresses, a given structure which was first represented within the interlanguage hypothesis via one or more mal-rules will be revised to eventually be represented by good grammar rules that correctly generate it. Note that before this happens, an initial revision may be in favor of other mal-rules representing misconceptions, but which are closer on the path to a correct characterization of the L2, the intended target of the sequence of interlanguage revisions.

Our presentation of a learner’s I_i can therefore be characterized as consisting of a set of language “rules,” each modeling how the language structures of the L2 are theorized by this learner at this point. Some rules will model standard L2 production, and some will model misconceptions. Since some of the structures involved are undergoing transition because they are in ZPD_i , there will also be competing rules, both correct and incorrect, which cover the same structure. When working out the nature of the L2 during Hypothesis Testing, the learner’s productions may vacillate between the competing realizations of a structure in the stage of currently being acquired.

One Grammar, Many Users

Since the learner generates sentences in the L2 using the rules which are in his or her interlanguage, it is clear that our parsing grammar must include the rules that could be found in any learner’s I_i in order to have the ability to recognize the syntax of any learner’s writing.

Two characteristics of ICICLE's user drive the determination of the contents of this parsing grammar. First, as mentioned earlier, ICICLE is designed to be used by a heterogeneous user group, spanning a broad range of English writing proficiency. Secondly, ICICLE is intended to follow a given user over time and the development of new language skills. Therefore, the parsing grammar cannot reflect a unique I_i but rather must reflect the union of all possible I_i . This means that it must contain all of those rules which could be in any I_i for any user of the system. This includes a complete, broad-coverage grammar representing the English language, to capture all of the structures which a learner may have already acquired. It also includes a broad range of mal-rules to cover any of the misconceptions a learner may have during the progress toward L2 competence.

Since this grammar is designed to ideally encompass all possible I_i , it is analogous to the concept of a "space" of possible student models mentioned by Sleeman (1982) and Opwis (1993), among others, usually defined, like in our case, as a set of rules and mal-rules. Our grammar can therefore be postulated as a *grammatical space* containing all possible interlanguages I_i . This space, which we will call I , contains:

- All correct grammatical rules which model standard L2 production.
- All possible incorrect rules modeling incorrect hypotheses of grammatical structures (mal-rules) exhibited by the population⁵.

Given this definition of I , a specific user's interlanguage state I_i at any point in time i would consist of a subset of I , specifically:

- Those correct grammatical rules which realize grammatical structures he or she has acquired.
- Those incorrect rules (mal-rules) which represent remaining misconceptions which the process of Hypothesis Testing has not yet repaired. This represents the user's implementation of those grammatical structures still unacquired.
- Both correct and incorrect rules covering those structures currently undergoing Hypothesis Testing (the current ZPD, which we will label ZPD_i).

This characterization of the user is illustrated in Figure 3. Those structures which are acquired are represented by correct rules (shaded squares), while those about which the user maintains misconceptions are shown as mal-rules (shaded "broken" shapes). Structures in this user's ZPD_i would be represented by both types of elements as different realizations compete with each other during Hypothesis Testing.

Reflecting the Individual

Our modeling task is therefore to determine what rule subset forms a user's I_i . We do this by observing his or her written productions in conjunction with an inferencing mechanism based on typical acquisition

⁵To develop a grammar to approximate I , we collected a large number of writing samples from the population. These samples spanned a wide range of writing proficiency and thus represented learners at various stages of acquisition. We have attempted to develop a grammar that could parse this corpus (Schneider & McCoy, 1998). The resulting grammar approximates I .

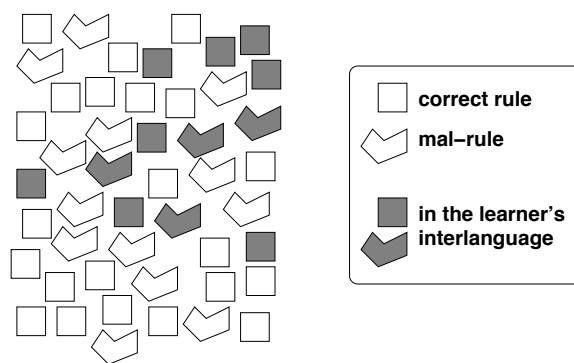


Fig.3. The user's interlanguage as a subset of grammatical space.

order in order to fill in missing knowledge⁶. Although the Interlanguage theory is based on verbal L2 production, observations by such researchers as Pennington (1996) have concluded that L2 writing is also reflective of a systematically changing cognitive interlanguage representation of the L2. Mangelsdorf (1989) holds that Hypothesis Testing is visible in written L2, and that learners performing Hypothesis Testing in a written format will also adjust their knowledge based on the feedback they receive, bringing their interlanguage closer to the L2. With respect to the L2 writing of deaf students, Russell *et al.* (1976) observe the existence of systematic rules which, while frequently representing a deviance from Standard English, approach a more standard model of the language over time.

Given this and other work in acquisition (c.f. Krashen, 1981; Krashen *et al.*, 1978; Kelly, 1987; Tarone, 1982), we can conclude that the kinds of expository texts which are the expected input to the ICICLE system will indeed reflect what the user has truly acquired, and therefore are much truer indicators of the learner's I_i than if the system were based on elicitation exercises, where the learner would be more focused on grammar than communication.

The implication is that we can expect structures which have been acquired previously to state I_i to occur in the learner's language usage without significant variation or error, while structures in the ZPD at state I_i (ZPD_i) will exhibit the variation typical of the transitional competence. Meanwhile, structures beyond the ZPD should either be absent from the learner's language production because of avoidance, or they should be used with consistent error because they cannot be avoided. At state I_{i+1} , many structures which were in ZPD_i have now been acquired and should be part of the learner's area of competence, and a new ZPD_{i+1} will have been formed as new rules come under the focus of the learner's Hypothesis Testing.

Implementation

For our implementation of the ICICLE user model, one of the approaches we have incorporated into our design is that of an *overlay* model. The underlying idea of an overlay is that the user's knowledge is

⁶This inferencing mechanism is discussed in more depth later.

represented as the subset of some knowledge space stored in the system⁷. Since we have described our user model as essentially representing the user's I_i , the knowledge space here involved is the space of grammatical constructs in the English language.

This overlay-based aspect of ICICLE's user modeling component is called MOGUL, for *Modeling Observed Grammar in the User's Language*. We refer to the basic unit of information in MOGUL as a Knowledge Unit, or KU, a term borrowed from Desmarais *et al.* (1996), who define the concept as representing a meaningful unit in knowledge of the domain, where the user's mastery of a KU can be reliably assessed. In the domain of language mastery, we see each language "structure" as a Knowledge Unit. They are essentially associated with bundles of the grammatical rules in our parsing grammar. For example, one KU might be Subject Relative Clauses. Associated with this KU would be all of the rules from the parsing grammar which implement this concept. This includes not only those rules modeling correct execution of this type of relative clause, but also the mal-rules which realize the ways in which this structure is executed incorrectly by the learner population. We assess the user's mastery of each KU as follows:

- Structures which the learner uses consistently correctly (reflecting grammar rules modeling standard English) are assessed as *acquired*.
- Structures which exhibit consistent error, when present — generated from mal-rules, presumably — are assessed as *unacquired*.
- Structures which exhibit variable performance are determined to be in the *ZPD*.

This mastery is recorded on each KU of the MOGUL model by tagging KUs as "acquired," "unacquired," or "ZPD." In the next section, we address how we have extended this basic idea to compensate for possibly incomplete data on the individual user.

Incomplete Knowledge

As described so far, the data stored in MOGUL is derived by the system from the written language performance of the user. An inescapable fact of deriving language data from freeform written text, however, is that through lack of opportunity or deliberate avoidance, many language structures will be absent (Corder, 1973; Wilbur *et al.*, 1976). This complicates our modeling process because we are forced to proceed on incomplete data about learner characteristics.

In the Grundy system, Rich (1979) also addressed the problem of basing system decisions on small amounts of possibly incomplete information about the user. In order to compensate for the fact that Grundy was able to collect data on only *some* characteristics of the user, Rich developed the idea of "stereotypes" which essentially represented populations of users and their associated typical characteristics. This allowed some of the data already collected on the user to "trigger" associated characteristics which could be inferred about the individual until such a time as other data contradicted that conclusion. An example in Grundy's domain (the suggestion of books to a library patron) would be the inference that

⁷Sometimes in an overlay model this subset is marked strictly in a boolean fashion — the user has acquired a particular unit, or has not — and sometimes degrees of certainty or degrees of mastery are indicated.

because a patron is female, she may be interested in a romantic novel. If the patron indicated otherwise during her interaction with Grundy, this inference would be overridden.

In our application we have been inspired by Rich's use of stereotypes. Analogous to the user characteristics she describes, we have the tags (acquired, unacquired, ZPD) marked on various KUs in the MOGUL model. However, we distinguish our approach from Rich in several key ways. We define our stereotypes not just as groups of users within a category but rather as collections of individuals who are generally at the same point on the path toward acquiring a language. Therefore, our collection of stereotypes is really a series of way-points a typical user passes through along this acquisition path. Rather than a specific characteristic of the user "triggering" a stereotype, the overall *similarity* of the user to a stereotype (i.e., looking at the collection of characteristics as a whole picture) is what causes stereotype data to be activated for that user (this will be discussed in more depth later). Finally, because of our definition of a stereotype, we are not merely concerned with enabling the system to *repair* incorrect stereotype selection to home in on a static user, but to allow for a user who is *learning*, so that the stereotype selection criteria must include adaptation to data which is changing as the user moves along the acquisition path to a new level.

One problem that we have faced is that if our stereotypes are to capture typical acquisition sequences, we must make explicit what that acquisition sequence is. While there has been some work in linguistics that indicates common patterns in L2 acquisition, no work has specified these patterns in sufficient detail to be able to capture them computationally. The next section addresses this problem.

The Order of Acquisition

In other user modeling systems lacking data about a specific KU with respect to a user, inferencing behavior is sometimes implemented to empower the system to resolve missing data on the user based on relationships between KUs about which the system has no knowledge (call these "empty") and those containing known tags. Typically, this is accomplished by modeling the acquisition relationships between the KUs. Often, the relationships indicated are *prerequisite* relationships; in other words, the model may show that KU_m (for example, algebra) is a prerequisite concept to understanding KU_n (for example, calculus). If, then, the user has exhibited a mastery of calculus, the system can *infer* that the user may have also mastered algebra even if there is no direct evidence for that conclusion.

In our domain, we can establish analogous relationships which exist in two dimensions: concurrent acquisition and order of acquisition. Researchers in L2 acquisition have suggested that the errors made by a language learner over time change in a *systematic* fashion (Dulay & Burt, 1974), and furthermore that there is support for a typical *sequence of acquisition* for language structures (Bailey *et al.*, 1974; Dulay & Burt, 1975; Gass, 1979; Krashen, 1982; Larsen-Freeman, 1976; Schwartz & Sprouse, 1996; Schwartz, 1998), sometimes called a "built-in syllabus" for L2 acquisition (Corder, 1967; Higgins, 1995). There are competing models of the acquisition process which account for this, and the details of the particular sequence of acquisition (i.e., the built-in syllabus itself) are rather vague in the literature. Without espousing a particular model, we borrow from the consensus of these accounts the notion of the acquisition of language occurring in some stereotypical sequence. We describe here our computational model for capturing this idea. We refer you to (Michaud & McCoy, 2004) for descriptions of how we

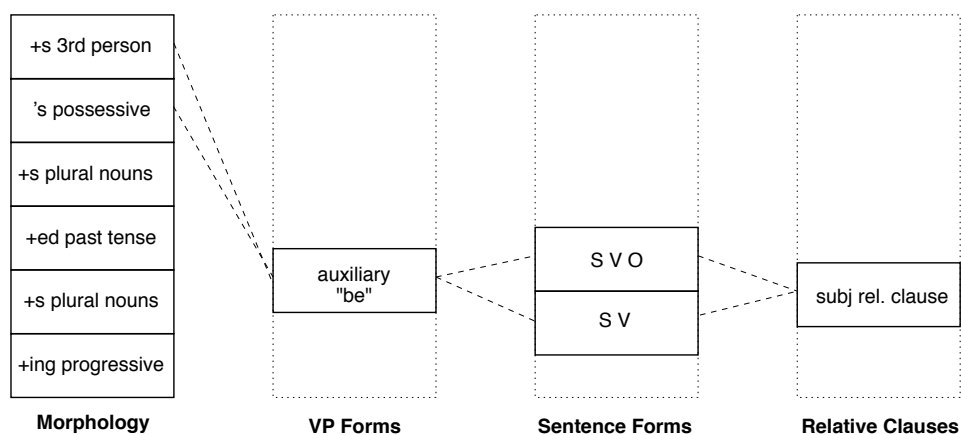


Fig.4. SLALOM: Steps of Language Acquisition in a Layered Organization Model.

have made empirical efforts involving a corpus study of written samples from users at various levels of proficiency in order to establish a concrete order for this syllabus.

We have incorporated the notion of a sequence of acquisition into an architecture which is based on a partial ordering of language structures (grouping together those structures which are acquired concurrently, and ordering these groups according to stereotypical sequences) in order to enable the ICICLE system to infer tags based on these acquisition relationships. Below, we describe how these acquisition orders are incorporated into our model of the user's interlanguage state.

SLALOM: An Organization on Grammatical Space

The name of the stereotype architecture which is coupled to the MOGUL model is SLALOM (*Steps of Language Acquisition in a Layered Organization Model*). The design was originally proposed in (McCoy *et al.*, 1996) and has significantly evolved over the years (Michaud & McCoy, 1999, 2000; Michaud *et al.*, 2001). A very simplified representation of SLALOM's basic structure can be found in Figure 4, where each KU is represented by a rectangular box.

The first half of SLALOM's name, *Steps of Language Acquisition*, refers to how SLALOM captures the stereotypic linear order of acquisition within certain "hierarchies"⁸ of morphological and/or syntactic structures such as negation, noun phrase construction, or relative clause formation. The figure depicts a hierarchy as a vertical stack of boxes. As an example, we illustrate a Morphology hierarchy based on the findings of Dulay & Burt (1975)⁹. Within these hierarchies, a given morphosyntactic KU is expected to be acquired subsequent to those below it, and prior to those above it, according to the natural order of a stereotypical learner from this particular L1 acquiring English¹⁰. The figure's example represents

⁸The term "hierarchy" refers to the fact that certain KUs are learned before others.

⁹This is for example purposes only, since their morphology sequence is relatively simple. The sequence we actually use can be found in (Michaud, 2002).

¹⁰Although some research has shown high correlation between the acquisition orders of learners from different L1s, we wish

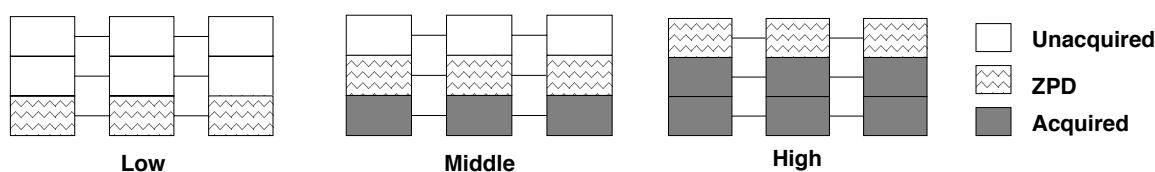


Fig.5. A simple view of stereotype progression in SLALOM.

the idea that “+ing progressive” is typically learned before “+s plural nouns,” which is typically learned before “+ed past tense.”

Since we wish to coordinate the acquisition stages between the hierarchies, dashed lateral connections in the figure represent the second dimension of relationship stored between the KUs in the model, namely that of concurrent acquisition¹¹. We call these lateral groupings “layers,” and the figure has one drawn in as an example¹². This is the source of the *Layered Organization Model* part of SLALOM’s name, referring to these layered groupings of KUs which essentially illustrate a progression through the acquisitional sequence representing those KUs being learned at certain stages of acquisition. In particular, we expect that students who are just beginning to learn written English will first struggle with those KUs in the “first layer” and then will progress “up” the hierarchies. A simple view of this progression can be seen in Figure 5. In reality, some KUs participate in more than one layer, which indicates the relative speed with which some concepts are acquired (i.e., KUs that span more than one layer take longer to master completely).

Like Rich’s stereotypes, SLALOM’s acquisition relationships allow the ICICLE system to infer absent data on the user from that which has been recorded. The uniqueness of the SLALOM concept is that it captures a sequence of stereotypes, not as individual images, but implicitly within one interlinked architecture. Its layers each indicate the structures which may at one stage in the learning process form part of that stage’s stereotypical ZPD. Those layers located “below” in the model contain structures which are typically mastered before, while those layers “above” are what is typically acquired later in the learning process, at a later stereotypical stage. If a KU which is untagged in MOGUL represents some language structure which SLALOM shows to be “below” an Acquired structure or a ZPD structure, then the system can infer that the untagged KU is probably also Acquired, even if it has not occurred previously in the user’s observed language production. Likewise, structures sharing the same layer in SLALOM should share the same tags in MOGUL, and structures “above” the ZPD layer can be inferred as Unacquired. These indirect conclusions are not as strong as those based on actual empirical data from the specific user, but in the absence of stronger data they can be used to make planning decisions.

In order to represent the acquisition sequence in this way, we must “place” each KU into SLALOM in such a way that illustrates its relationships to other KUs according to how a typical learner moves

to represent the most likely order possible and thus have restricted the model to representing learners from a specific L1, in our case ASL.

¹¹This relationship may also be indicated within a hierarchy, not shown in this figure, to capture a partial ordering in which some structures in the same hierarchy are acquired together.

¹²The contents of this layer are not based on empirical findings and are for illustration only.

through the acquisition sequence. A prototype placement has been implemented, based on an investigation of acquisition data among deaf students summarized in (Quigley *et al.*, 1977) and (Wilbur, 1977). We again refer the reader to (Michaud & McCoy, 2004) for a discussion of our ongoing work to develop SLALOM placements based on our own corpus of writing data.

Representing a Changing User

Many user modeling systems are more concerned with how to utilize the information stored within the model in order to drive system decisions than with the tasks of initially building up that information about the user (c.f. Lesgold *et al.*, 1992; Woolf, 1984) or with changing it to reflect a dynamic user over time (c.f. Paris, 1988, 1993). In those systems, the user model is often static and assumed “given” from some outside source. Bull *et al.* (1995a) point out, however, that a *student* model must be dynamic in order to change as learning occurs. Since ICICLE’s user model needs to capture fine details about the user derived from user performance, and ICICLE will be used by an individual over time and across the development of new skills, it is central to the model design that it be established how this model is initialized and updated over the course of time.

Initialization

The initialization data for ICICLE’s user model comes from the first parses performed by the error identification component and the subsequent first analysis of the user’s performance on the KUs captured in the system. Once this information is received by the user modeling component, it can assign tags in the model according to how the user performed on his or her first writing sample and all of the language structures it contains.

One serious issue this raises is that the first parse of the first piece of writing by any new user must be performed without the use of the user model to help the system select between parses. This appears to be a non-trivial problem, involving the need for information on *expected* user performance without yet having any data on *exhibited* user performance. One approach taken by similar systems such as EDGE (Cawsey, 1993) is to base the initial state of its user model on a general classification of the user provided to the system. We have embraced a similar approach; a new user account is created with a classification of “Low,” “Middle,” or “High” written English proficiency, depending on the user’s self-reported classification. This is a *stereotype* classification only, essentially placing the user’s current acquisition level at one of the layers in SLALOM; it is not recorded within the MOGUL model, because it is not considered to be as reliable as genuine grammatical competency data specific to the individual. In the domain of language and writing proficiency, it is difficult to obtain self-assigned or even teacher-assigned classifications which are not highly subjective and variable according to individual standards. This classification suffices, however, to provide a “best guess” framework on which to base the first analysis of a new user’s work¹³.

¹³For our discussion of an evaluation in which we test how well ICICLE adapts from an initial stereotype guess to a more accurate portrayal of the user, please see (Michaud *et al.*, 2005).

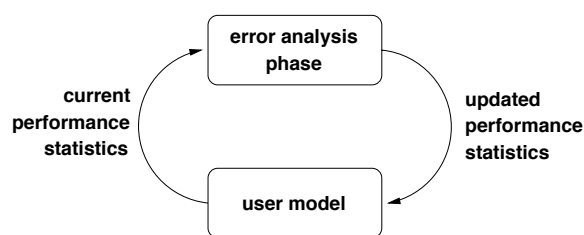


Fig.6. The retrieval/update cycle between the error analysis phase and the user model.

Once ICICLE successfully performs this first analysis, an initial set of tags is placed on the KUs representing language structures in MOGUL. This information immediately takes precedence over the initial SLALOM placement and determines any future relationship between the user and any assumed stereotype, because it is our belief that the MOGUL model, being based directly on the user's writing in multi-sentence samples of potentially significant length, has a very rich and reliable source of input on the user's actual language proficiency. In comparison to polling user knowledge where one question is only likely to reveal one point of data (either the user understands or does not understand the concept asked about), even a short piece of writing is going to offer many points of data per utterance. Every grammatical construct successfully or unsuccessfully used, from determiner choice to verb tense, provides information about the user. These points can be correlated to provide a map of those constructs used correctly, those which are experiencing variation, those which are occurring with error, and those which are absent; therefore, even after only a single writing sample, the MOGUL profile of the user's interlanguage will contain a rich store of directly-derived data about this individual. This avoids the sometimes undesirable activity (Anderson *et al.*, 1990; Rich, 1979) of having to question the user extensively in order to determine appropriate system actions. Finally, as discussed above, free-form writing for communicative purposes is a more reliable source of interlanguage data than grammar performance elicited through tasks such as translation or fill-in-the-blank exercises.

Updating

Every time ICICLE performs an analysis of user text, the system has new (and potentially different) observations of user performance to add to MOGUL, and that information is sent along so that it may cause the model to be updated accordingly. The relationship of the user model to the error analysis phase is represented in Figure 6. The tags in MOGUL may be deemed incorrect for one of two reasons: an existing tag might have been based on a small sample size of user performance and more data might show that it is wrong, or the user's proficiency may simply be changing, as we hope it will over time. In either case, a tag may possibly need to be revised either "upward" (e.g., from "ZPD" to "Acquired") or "downward" (e.g., from "Acquired" to "ZPD"). ICICLE must therefore have a mechanism for re-evaluating these tags and updating their values.

Glaser *et al.* (1987) note that learning assessment must be able to account for changes in performance as new stages of acquisition are manifested in learner performance. However, a model that can be overwritten over time gives rise to the question of whether new data should always champion over

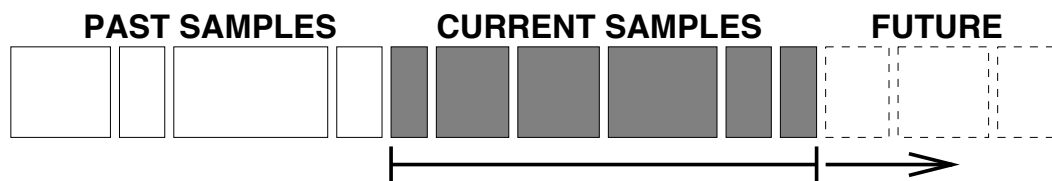


Fig.7. The “sliding window” of performance history.

the old, as with the EDGE system, which always overwrites previously recorded tags (Cawsey, 1993). The guidelines given thus far of how observations are recorded in the model have been that features used “consistently correctly” are acquired concepts, those used “with variation” are in the ZPD, and those which appear “consistently incorrectly” have not yet been acquired. As the system goes through more than one piece of the student’s work and the amount of data increases, these judgments may change, particularly if one or more of the pieces is too short to contain several instances of certain language structures, or a structure is simply not a commonly used one. Therefore, it makes sense for the model to track certain figures (the number of times a structure has been attempted, and the number of times it was 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 allows 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, our model architecture includes the maintenance of a “sliding window” of performance history across the continuum of writing samples stretching into the past from the current sample (see Figure 7). Ideally, this window includes enough data to be robust, and yet be small enough to capture only the “current” statistics¹⁴. This latter requirement is particularly important for the system’s self-evaluation and deciding whether recent explanatory attempts have succeeded.

One advantage to using the “sliding window” approach is that the user model always reflects the user’s “current” levels of proficiency, even if that proficiency changes in a *backward* direction. It is, of course, our hope that backsliding does not occur in our users, but in some cases it has been observed that a language learner who has been performing well on a structure will start to make errors again at a later time. By always tracking how the user is performing only in the “recent” past, the user modeling component will be able to revise its tags in either direction as that performance changes.

There are several possible reasons to account for the reintroduction of errors in performance, including the over-generalization of rules, the discarding of unanalyzed formulae which formerly produced

¹⁴We have not yet implemented the sliding window in the prototype system, because at this time we have not had the opportunity to collect significant longitudinal data or to deploy the system with an actual user over a period of time, as we intend for the future use of the system.

correct performance without the underlying grammatical knowledge, or the effect of the cognitive load created by learning new concepts or focusing more on discourse-building tasks. Although some of these situations represent a backsliding in performance while actual proficiency may be advancing, it should be noted that the objective of MOGUL is to represent a profile of the user's current *performance* in order to select the most appropriate parses to match. Therefore, a downward revision of a tag would be an accurate reflection of the current production of errors, even if conceptually it seems inappropriate to tag a structure as being Unacquired or in the ZPD if the learner is actually very close to full mastery. The tag will catch up to the learner's actual proficiency when the performance does¹⁵.

APPLICATIONS OF THE MODEL

The information in the user model supplies two kinds of data to the various ICICLE processes: direct data on the user's performance from MOGUL, and indirect data which can be inferred at need from the relationships of untagged items to tagged ones in the SLALOM hierarchies. The latter data will be less reliable, and ICICLE will take note of this by weakening the conclusions drawn from it. Details on how the user model is and will be applied are discussed below.

Parse Disambiguation: Choosing a Parse Using Likely Rules

As discussed earlier, a given learner from our population generates written English from some I_i , or a subset of the parsing grammar which includes some standard rules and some mal-rules depending on his or her level of development. Our MOGUL model captures this I_i by indicating which grammatical structures are acquired (and thus are realized in I_i by standard rules), unacquired (realized in I_i by mal-rules), and in the ZPD (realized in I_i by competing rules which result in variable performance).

This information is used by the Parse Tree Selector module of ICICLE to select a parse tree that uses rules that are most likely to be in I_i . Recall that a parse tree of a sentence is made up of rules in the grammar that were used to parse the sentence. In the current implementation, we use both portions of our user model (the direct proficiency data in MOGUL, and the stereotypic acquisition data in SLALOM) to score each parse tree with the intent of giving trees that are made of rules that are believed to be in the user's I_i higher scores than those which are constructed of rules that are not likely to be in the user's I_i (and therefore unlikely to appear in their language production).

The scores are assigned as shown in Table 1. If a grammatical structure has not been acquired by the user, we do not expect that user to be able to execute it without error; therefore, a tree node formed by a mal-rule representing that structure is scored **+1**, while a node formed by a regular (correct) rule is scored **-1**. In the case of an acquired structure, the expectation is reversed — that the user would be likely to use it correctly, not commit an error while attempting it. Therefore, nodes formed by regular rules are scored high (**+1**) and nodes formed by mal-rules are scored low (**-1**). Structures in a user's ZPD

¹⁵We do acknowledge that a problem with 'inaccurate' tags from a proficiency point of view is how the user modeling component may use SLALOM to draw incorrect inferences from these underestimated tags. This is another of the reasons why inferences based on the relationships in SLALOM are not as certain as those based directly on actual performance from the learner.

Table 1
Calculating Node Scores.

MOGUL Tag	Rule Type	In I_i?	Score
Unacquired	Regular	no	-1
	Mal-Rule	YES	1
ZPD	Regular	YES	1
	Mal-Rule	YES	1
Acquired	Regular	YES	1
	Mal-Rule	no	-1

are realized by competing correct and incorrect rules in I_i , so ZPD-marked structures are scored the same regardless of whether the structure is executed correctly or poorly.

Given this assignment of scores throughout the nodes of the trees, each parse tree is assigned a single score which is the average score of all of the rules used in the parse tree. A tree which has the highest score¹⁶ is selected as the single most likely parse.

One of the benefits of selecting a [-1, 1] interval over which to express these scores is the notion of 1 representing a “strong positive” and -1 a “strong negative.” The value 0, being in between, would reflect an inability to state a belief about the rule’s status. Although 0 is never assigned to a rule node¹⁷, it is clear that later iterations of the scoring mechanism may use the strength of evidence on a tag to produce a measurement of the *confidence* in the score. If this confidence measure were expressed in terms of a percentage, e.g. 90%, it could be applied against either “yes” or “no” scores by multiplication to affect the strength of the score. Low confidence levels would bring the score closer to 0, the neutral statement.

We consider this to be a reasonable approach toward parse disambiguation given that true intelligent disambiguation requires recognition of learner intent (Corder, 1974); in the absence of this, we have only what Corder labeled “plausible interpretations.” However, with this approach to a kind of “probabilistic” grammar (where rules in I_i are considered more likely), ICICLE has access to information determining which of these interpretations is more plausible than the others. This also addresses the problem raised by Sleeman (1982) and Self (1988) of the immense search space that is daunting in this kind of user modeling by quickly giving access to the most likely interpretation.

Focusing Instruction on the Frontier of Acquisition

We discussed earlier the importance of focusing generated tutorial text on the frontier of the learner’s acquisition process, which was later equated to those structures in ZPD_i . As established above, instruction and corrective feedback on aspects of the knowledge within ZPD_i may be beneficial, while instruction dealing with that outside of the Zone is likely to be ineffective or even detrimental. When passing the error list to the response module, therefore, the error identification module of ICICLE will consult the

¹⁶At this time there may be more than one parse tree receiving the highest score; we are working to further strengthen the differentiation power of the scoring procedure, as discussed in (Michaud *et al.*, 2005).

¹⁷Except in a small number of situations where a rule in the parsing grammar does not represent a grammatical structure, as is the case with some “fix-it” rules which merely insert certain key features. These nodes are not included in the tree average.

MOGUL tags described here in order to prune the errors so that the tutorial responses are focused only on those errors at the user's current level of language acquisition (in ZPD_i).

EVALUATING THE MODEL

As mentioned above, ICICLE's user model has been implemented in the system with a MOGUL component and a prototype SLALOM architecture representing stereotypical acquisition sequences. We have performed several evaluations of the performance of the model and have published the complete discussion of those evaluations and their results in (Michaud *et al.*, 2005). Here we briefly summarize the results.

Currently the ICICLE system is a prototype implementation that has only been tested in the laboratory by researchers using testing data from our aforementioned collection of writing samples from students who are deaf. We acknowledge that in order to evaluate whether or not the ICICLE system's approach to language instruction is beneficial to a "real user", much additional work would need to be done to ensure system robustness in the face of naive users with unrestricted text input. Given the limits of the prototype, however, theoretical testing has illustrated that our approach to user modeling shows significant promise in parse disambiguation, adaptation to the changing learner, and personalization to a non-stereotypic acquisition journey.

Parse Selection Based on the User Model

We examined the ICICLE system's parse selection process to see if it successfully makes "correct"¹⁸ parse selections based on the user model. We provided the system with 15 randomly-selected sentences to parse, and tested these sentences on each of three hypothetical, stereotype-based users: a Low-level learner, a Middle-level learner, and a High-level learner. We examined the parse selection choices for each level, which in each case accurately reflected the stereotypic performance expected by users at each of those phases in acquisition, rewarding those interpretations containing structures we expect to be in I_i for a learner of that level, while penalizing those we do not believe to be in I_i .

Following the User's Learning Journey

In this evaluation, we provided ICICLE with an initial user stereotype placement, and then provided the system with a series of utterances representing higher performance than the stereotype placement indicated. By doing this, we sought to simulate a user whose proficiency was changing in a positive fashion, in order to see if the system would recognize this difference and update the model accordingly¹⁹. Recall that we expect our user's language proficiency to be dynamic as learning progresses, and that eventually the user model record for a user at step i in his or her language acquisition will adapt when the user is at some later step $> i$.

¹⁸As compared to a human judge.

¹⁹There is no bias toward upward revision of the user's learning status. Although we chose to illustrate a learner's upward transition in this example, the ability of the user modeling component to adjust the stereotype is the same whether it is being revised higher or lower.

We selected a batch of 20 sentences from samples in our corpus, all representing the successful execution of higher-level structures that would be expected from the user as described in the initial stereotype placement. The system successfully recognized the improvement in user performance and updated the user model information appropriately; in fact, when given the sentences in small increments, allowing for a user model update between each utterance, this update was achieved before the end of the 20 sentences and resulted in a slightly higher level of parse selection accuracy. The original run had agreed with a human judge on parse selection for 12/20 (60%) of the sentences, while the sequential run's agreement rose to 15/20 (75%) of the sentences, an effect that was obviously concentrated on the last few sentences processed as by that point the system had at its disposal new evidence on the user's proficiency level.

The Unique User

While some users may fit stereotypes well, each learner is unique, and may have a unique set of mastered skills and skills which he or she has yet to learn. In a final evaluation task, we explored the integration of the two facets of the user model; specifically, we tested ICICLE's ability to integrate the stereotype expectations which SLALOM provides with the specific history of parses that it records in the MOGUL model.

For this task, we chose to create a profile representing an advanced learner in whose language mastery there exists certain fossilized structures that are executed with error in a fashion "atypical" of that level of performance²⁰. We sought to illustrate that if the system has built up a performance history for this user illustrating these differences, it will select parses more appropriate to the user's actual interlanguage I_i even if that deviates from stereotype expectations.

For this evaluation, we chose from the corpus 35 example sentences to represent our learner with fossilized errors. These sentences exhibited advanced constructions, but many (17/35) contained errors in some basic elements normally acquired by an advanced learner. We trained the MOGUL model on portions of this selection of sentences and then tested the recorded user data on held-back sentences to see if the uniqueness of this user's language mastery was accurately reflected in MOGUL. The stereotype expectations of SLALOM would not have always indicated the correct parse because of the decreased likelihood of such "beginner" errors; in fact, an earlier iteration of this evaluation discussed in (Michaud, 2002) indicated that the stereotype alone only agreed with the human judge on sentences which were "typical" of the stereotype, and failed to identify the correct parse on 50% of the sentences containing low-level errors. However, with the stereotype profile augmented by the MOGUL profile of the user's individual performance, the system was able to perform much better in the parse selection task, and correctly selected the same parse as a human judge on every one of the test sentences.

²⁰Fossilization is actually a common occurrence in second language learning, reflecting some low-level linguistic elements which remain unacquired even after the learner has significantly progressed.

COMPARISON TO OTHER SYSTEMS

Systems for the Deaf

The use of computers with the deaf has a long history, stretching back to the beginning of the 1960s (c.f. Richardson, 1981, for a survey); as early as 1968, computers with writing programs were being used at NTID²¹ to assist the development of English syntax (Watson, 1979). Overviewed here are two of ICICLE's peers in the domain of computer systems for the deaf in order to contrast their approaches with ours.

Ms. Plurabelle

The Ms. Plurabelle grammar-checker (Loritz *et al.*, 1993; Loritz & Zambrano, 1990) was developed in conjunction with Gallaudet University (the only liberal arts college for the deaf) with the intention of creating a system capable of handling the written English of deaf students with a high degree of accuracy. Like ICICLE, it consisted of a parser and a feedback mechanism which pointed out errors to the user. Unlike ICICLE, it did nothing to model the individual user.

Although the creators of Ms. Plurabelle did agree with the goal of avoiding pointing out errors beyond the ability of the user to correct (Loritz & Zambrano, 1990), they did not avoid instructing on concepts well-mastered by the student as discussed earlier. Instead, they simply worked from an initialized proficiency level representing the student (a numerical value between 1.0 and 5.9) which allowed the system to prune error notifications which it judged would be beyond the student's understanding. This proficiency level was static and was provided to the system at onset of use.

In testing with actual users, it was found that even the pruned error messages were so confusing to the users that they got themselves mired in decoding the message content²². Users were actually more successful at determining the correct version of the sentence if they were simply allowed to experiment with alternative hypotheses without any feedback from the system (aside from flagging sentences with errors).

Loritz & Zambrano (1990) admit there are many limitations to the Ms. Plurabelle system, including the fact that it stops parsing at the first error in the sentence (and is therefore unable to report on more than one error per sentence). For the domain of parsing the English of deaf students, however, it was a clear improvement over commercial grammar-checkers because the parser (ENGPARS) was a generalized transition network trained on a corpus of writing from the population and therefore it performed at a much higher level of accuracy than systems which were designed for hearing individuals writing in their native language. ENGPARS was reported to reach as high as 90% accuracy. Within the context of a grammar-checker, this performance improvement is significant and to be appreciated; we acknowledge that ICICLE's parser has not yet been similarly tested with its real user population. However, within the context of assisting students in improving their grammatical performance, adapting to individual students, or changing to meet the needs of a student over the learning process, Ms. Plurabelle's approach

²¹The National Technical Institute for the Deaf.

²²This motivates one of our current directions in the ICICLE project, the implementation of an animated agent on the screen that can use some signed communicate to express some of the feedback to the user.

would fall far short. Although the system was trained to parse the writing of a population of users, it is a general population, and the approach fails to account for a heterogeneous group of learners whose different levels of grammatical competence might result in different parses being more likely. This type of adaptability is the core motivation behind our work with the ICICLE system.

Writing Safari

The “Writing Safari” system (Farrow *et al.*, 1994) presents a somewhat complementary approach to that discussed above. Where Ms. Plurabelle concentrated on enabling a system to parse the grammatical productions of the deaf population without providing instruction to the students, Writing Safari is more of a learning environment than a parser, presenting to the user a kind of back-plot to the writing process, inviting him to accompany a set of characters on a choice of trips to Kadaku and exposing him to samples of writing from these characters in a variety of genres — journal entries, letters home, etc. — the selection of which depends on the trip chosen. It leaves the actual analysis of the writing in the hands of peers and teachers. The instructional areas of the system are independently navigated by the user; for instance, a special “danger zone” area of the knowledge presented on the system warns the writer of “deafisms”, or common mistakes made in these contexts.

Although Writing Safari in general represents a very different kind of environment and emphasis from that envisioned for ICICLE, the philosophies of the authors are not incompatible with ours. They stress that errors should be viewed as “a reflection of a developing understanding of how English grammar is used to express a certain set of ideas”, a view very much in tune with our approach to using user performance as a key to describing his or her developing language proficiency. Like the creators of Ms. Plurabelle, Farrow *et al.* note that most grammar correction is wasted because the receiving student is not developmentally ready for it. They also mention that the ZPD theory illustrates the importance and efficacy of adjusting instruction to the student.

Discussion

A survey conducted by Rose & Waldron (1984) showed an explosion of computer use in programs for hearing-impaired and deaf students in the mid-80s; McNally *et al.* (1987) lamented, however, that while the use of computers with deaf children at that time was becoming very popular, the rising popularity had “outdistanced” the understanding of how computerized systems could help deaf students develop written language skills. It is our hope that ICICLE can prove this to no longer be the case.

The direction ICICLE takes with respect to deaf education could be seen as sitting somewhere between the two approaches embodied by the Ms. Plurabelle and Writing Safari systems. We wish to capture the automated grammatical analysis that formed the core of the Ms. Plurabelle system, while providing instruction and an encouraging writing environment like Writing Safari. A top-down, communication-centered instructional emphasis like that modeled in Writing Safari could be very easily used in the classroom in conjunction with the syntactic feedback from ICICLE. Although Farrow *et al.* argue that grammar-checkers make deaf students over-reliant on the computer’s ability to correct their

errors²³, and one could see that becoming the case with students making use of a Ms. Plurabelle-style system, ICICLE's focus on an attempt to *instruct* the user — and, presumably, provoke him to correct the sentence on his own without the system simply providing the answer — provides less of a crutch and more of an encouragement to learn the reasons behind the error well enough to avoid repetition.

Other Language Learning Systems

This section briefly discusses other CALL systems which are comparable to ICICLE and how they have addressed similar challenges to the ones ICICLE faces.

BELLOC

The designers of BELLOC (Chanier *et al.*, 1992) share with us the approach of viewing the utterances produced by users as representative of sets of internal rules modeling language. In their application for English-speaking learners of French, users interact in written French within the domain of negotiating an inheritance. When a user produces an error, BELLOC seeks to determine the underlying rule (what we would refer to as a “mal-rule,” and Chanier *et al.* refer to as an “applicable rule” or AR) in order to provide constructive and helpful feedback to the learner.

Like in ICICLE, their grammar is augmented by a set of typical “learner’s rules,” or error-containing rules from their population. They do not, however, take note of which of these rules (if any) a given user can be expected to use. If correct rules fail to parse a sentence, a theorem-prover attempts to select from the set of error-containing rules to achieve a parse. For each sentence it cannot parse with correct rules, the system narrows the candidate applicable rules down to a specific set which could explain the resultant utterance, and then explicitly queries the user on judgments of certain error-containing phrases to determine which one the user believes is acceptable French, using this selection to diagnose which underlying rule is causing the current problem. This extra negotiation with the user would clearly be overly intrusive if the system were interpreting an extended piece of text, particularly if the user produced several errors per utterance.

HyperTutor

The HyperTutor system (Schuster & Burckett-Picker, 1996) is a learning tool for Spanish-speaking English learners of reasonable proficiency. It interacts with the student through a series of translation tasks, presenting Spanish sentences which the student then translates into English. Like ICICLE, the system gives notification on correctness and explanations of the error(s) found, if any.

Although the authors characterize the HyperTutor user model as modeling the learner’s interlanguage, their approach is significantly different from ours. The essential nature of their model is a store of the *language learning strategies* the system has observed the student using, the idea being that a learner applies these strategies to build the interlanguage and to provoke transitions. Possible strategies include applying the L1 grammar to the L2 or over-generalizing an L2 construct to a larger set of instances than

²³The authors of this paper would argue that they do the same for hearing students.

is appropriate. The HyperTutor consists therefore not of an actual profile of the interlanguage I_i , but rather of what possible reasons may lie behind an incorrect rule existing in I_i . While this supplies useful information to a tutorial component, it would fall far short of enabling the system to handle interpretation tasks outside of the prescribed domain of phrase translation.

German Tutor

German Tutor (Heift & McFetridge, 1999), a CALL system for English-speaking learners of German, accepts single sentences, parses them, and provides the user with feedback on the most salient error found. Its student modeling architecture is very similar to MOGUL; it utilizes a database containing all of the grammatical “constraints” the parser can recognize as met or broken (analogous to our KUs), each holding a score from 0 to 30 representing the user’s knowledge on this constraint. A score from 0-9 represents expert knowledge, 10-20 is intermediate, and 21-30 is novice. As the system records user performance over time, these scores are incremented with each observed failure to execute a constraint correctly, and decremented with each success.

The details of this user model, however, are lost in the parse selection process, where the student’s proficiency scores on each constraint in the user model are averaged, yielding a general proficiency score for the student. The possible parses are then ordered from simple to difficult, and the student’s general proficiency level is used as an index into that list. This technique fails to take advantage of the information about the student’s individual strengths and weaknesses that was stored in the model.

Mr. Collins

The CALL system Mr. Collins²⁴ (COLLaboratively constructed, INSpectable student model) (Bull, 1994, 1997; Bull *et al.*, 1995a,b) is perhaps ICICLE’s closest kin in the field of CALL, although its domain is restricted to the acquisition of Portuguese clitic pronoun usage, and its interaction with the user is largely drill-and-test, missing-constituent (fill-in-the-gap) questions. Mr. Collins’ instructional objective is also largely concerned with the learning strategies being employed by the learner.

Relevant to this work, however, is the fact that Mr. Collins models the dynamic user through a *sequence* of student models illustrating the acquisition order in the domain of Portuguese clitic pronoun usage, spanning a spectrum from the novice to the expert (a native Portuguese speaker). Like in ICICLE’s user model, Mr. Collins combines this acquisition information with direct information about the user. When attempting to parse a sentence, an “expert model” is consulted first — in order to try to parse the sentence as grammatical — and if this fails, the current model of user’s misconceptions is attempted. If this fails as well, the system goes further back in the user’s path along the acquisition sequence to attempt a parse with errors from earlier in the history. Subsequent to these attempts, the last place consulted is the realm of grammatical transfers from other languages which the student has learned.

²⁴“Mr. Collins” actually refers to the user modeling component of a larger system, but the name is also used for the entire system for simplicity.

Discussion

The majority of the systems discussed here have far more narrow goals than ICICLE, primarily addressing very specific aspects of the L2 within strictly-defined exercises such as translation. This strongly affects their user modeling requirements and objectives. BELLOC operates within a restricted domain, HyperTutor only parses strict translations of sentences it provides, and Mr. Collins instructs only upon specific syntactic categories and not the broader grammar of the language itself. Of the systems reviewed, only German Tutor accepts free-form text as ICICLE does. It is a very ambitious undertaking and it is clear from our work so far that ICICLE has some growing yet to do before it can consider this task truly conquered.

Perhaps because ICICLE is such an ambitious project, our user modeling effort is far more precise than those others discussed in this paper. BELLOC does not track the user at all; it has to choose between competing interpretations through directly querying the student, a time-consuming and intrusive task when there are many user utterances and/or many errors per utterance. Although HyperTutor shares ICICLE's goal of interpreting user actions through a view of the interlanguage, it is clear that Hypertutor's portrayal of a collection of learner strategies gives a very coarse-grained view of what learner rules may be in that interlanguage. It proposes to offer theories about which types of hypotheses may exist there without giving specific information about which hypotheses *are* there. HyperTutor also does not address the possibility of ambiguous errors (where more than one parse tree, and therefore more than one "cause" of the error, could account for a user's utterance), which is a key component to the issues we face regularly with the ICICLE system.

German Tutor's user modeling technique is similar to the MOGUL facet of our user model, but in their application they lose the precision represented by tracking user performance on each individual structure when they translate it into a global competence level. This approach completely ignores the usefulness of knowing that an individual may exhibit competence or preference for specific structures while he or she struggles on others, and it relies heavily on the ability to "rank" potential parses from most complicated to least. It also fails to engage any kind of stereotypic inference structure as is embodied in SLALOM.

Finally, while Mr. Collins' approach is remarkably similar to ours, it captures such a tiny piece of the second language acquisition question — only pronouns, and a subset of pronoun usage at that — that it does not face many of the complexities that we have addressed in our user modeling component for ICICLE.

It is therefore possible to conclude that the ICICLE approach to user modeling through the MOGUL and SLALOM facets, while incorporating some aspects of user modeling and user interpretation which have been seen before, is nonetheless novel both in the precision of concept and application of its user modeling component and in the ambitious scope of the entire story of syntactic acquisition which is embodied within the model.

CONCLUSION

In this paper, we have discussed how we address the adaptivity needs of a CALL system for deaf learners of written English by modeling the user's current interlanguage grammar. This enables our system, ICICLE, to intelligently address the problem of natural language parse disambiguation, and provides strong evidence to direct selective tutorial efforts toward those domain elements on the edge of acquisition. We have presented the architecture of a user modeling component that allows us to infer the user's interlanguage on the basis of his or her language production recorded in the MOGUL component of the model. The user modeling component supplements this performance data with knowledge derived from the SLALOM architecture which captures the stereotypical precedence and concurrence of acquisition among grammatical structures. The SLALOM architecture is used like a stereotype to infer a more complete image of user grammatical competence than is directly visible in his or her language productions. Our end goal is an intelligent tutoring system with the ability to respond appropriately to student learning difficulties, adjusting to the individual as it interacts with him or her across the learning journey.

ACKNOWLEDGEMENTS

This work has been supported by NSF Grants #GER-9354869 and #IIS-9978021. We would like to thank the ICICLE group at the University of Delaware, in particular: Rashida Davis, Christopher Pennington, Ezra Kissel, David Derman, H. Gregory Silber, and Michael Bloodgood for their work on the ICICLE implementation.

REFERENCES

- Allen, J. (1995). *Natural Language Understanding*. Second edition. California: Benjamin/Cummings.
- Anderson, J. R., Boyle, C. F., Corbett, A. T., & Lewis, M. W. (1990). Cognitive modeling and intelligent tutoring. *Artificial Intelligence*, 42(1), 7–49.
- Antworth, E. L. (1990). *PC-KIMMO: A two-level processor for morphological analysis*. Number 16 in Occasional Publications in Academic Computing. Summer Institute of Linguistics, Dallas, TX.
- Bailey, N., Madden, C., & Krashen, S. D. (1974). Is there a 'natural sequence' in adult second language learning? *Language Learning*, 24(2), 235–243.
- Baker, C., & Cokely, D. (1980). *American Sign Language: A Teacher's Resource Text on Grammar and Culture*. Silver Spring, MD: TJ Publishers.
- Barnum, M. (1984). In support of bilingual/bicultural education for deaf children. *American Annals of the Deaf*, 129, 404–408.
- Bialystok, E. (1978). A theoretical model of second language learning. *Language Learning*, 28(1), 69–83.
- Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.) *Cognition and the Development of Language*, Chapter 1, (pp. 11–54). New York: John Wiley & Sons.
- Bull, S. (1994). Student modelling for second language acquisition. *Computers & Education*, 23(1/2), 13–20.
- Bull, S. (1997). Promoting effective learning strategy use in CALL. *Computer Assisted Language Learning*, 10(1), 3–39.
- Bull, S., Brna, P., & Pain, H. (1995a). Extending the scope of the student model. *User Modeling and User-Adapted Interaction*, 5(1), 45–65.

- Bull, S., Pain, H., & Brna, P. (1995b). Mr. Collins: a collaboratively constructed, inspectable student model for intelligent computer assisted language learning. *Instructional Science*, 23(1-3), 65–87.
- Carroll, S. E. (1995). The irrelevance of verbal feedback to language learning. In L. Eubank, L. Selinker & M. Sharwood Smith (Eds.) *The Current State of Interlanguage: Studies in Honor of William E. Rutherford* (pp.73–88). Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Cawsey, A. (1993). *Explanation and Interaction: The Computer Generation of Explanatory Dialogues*. Cambridge, MA: MIT Press.
- Chanier, T., Pengelly, M., Twidale, M., & Self, J. (1992). Conceptual modelling in error analysis in computer-assisted language learning systems. In M. L. Swartz & M. Yazdani (Eds.) *Intelligent Tutoring Systems for Second-Language Learning* (pp. 125–150). Vol F80 of *NATO ASI Series*. Berlin Heidelberg: Springer-Verlag.
- Charrow, V. R., & Fletcher, J. D. (1974). English as the second language of deaf children. *Developmental Psychology*, 10(4), 463–470.
- Charrow, V. R., & Wilbur, R. B. (1975). The deaf child as a linguistic minority. *Theory into Practice*, 14(5), 353–359.
- Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, 5(4), 161–170.
- Corder, S. P. (1973). The elicitation of interlanguage. In G. Nickel (Ed.) *Special Issue of IRAL on the Occasion of Bertil Malmberg's 60th Birthday* (pp. 51–63). Heidelberg, Germany: Julius Groos Verlag. International Review of Applied Linguistics.
- Corder, S. P. (1974). Error analysis. In J. P. B. Allen & S. P. Corder (Eds.) *Techniques in Applied Linguistics*. Vol 3 of *The Edinburgh Course in Applied Linguistics*, Chapter 5 (pp. 122–154). Oxford University Press.
- Desmarais, M. C., Maluf, A., & Jiu, J. (1996). User-expertise modeling with empirically derived probabilistic implication networks. *User Modeling and User-Adapted Interaction*, 5(3/4), 283–315.
- Drasgow, E. (1993). Bilingual/bicultural deaf education: An overview. *Sign Language Studies*, 80, 243–266.
- Dulay, H. C., & Burt, M. K. (1974). Errors and strategies in child second language acquisition. *TESOL Quarterly*, 8(2), 129–136.
- Dulay, H. C., & 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.
- Erting, C. (1978). Language policy and deaf ethnicity in the United States. *Sign Language Studies*, 19, 139–152.
- Farrow, K., Power, D., & Freebody, P. (1994). Computer-assisted writing development for deaf students: "Writing Safari". *On-CALL Online*, 9(1). <http://www.cltr.uq.edu.au/oncall/farrow91.html>.
- Gass, S. (1979). Language transfer and universal grammatical relations. *Language Learning*, 29(2), 327–344.
- Glaser, R., Lesgold, A., & Lajoie, S. (1987). Toward a cognitive theory for the measurement of achievement. In R. R. Ronning, J. A. Glover, J. C. Conoley, & J. C. Witt (Eds.) *The Influence of Cognitive Psychology on Testing*, Vol 3 of *Buros-Nebraska Symposium on Measurement and Testing*, Chapter 3, (pp. 41–85). New Jersey: Lawrence Erlbaum Associates.
- Grishman, R., Macleod, C., & Meyers, A. (1994). Complex syntax: Building a computational lexicon. In *Proceedings of the 15th International Conference on Computational Linguistics*, Kyoto, Japan. Coling '94.
- Gutiérrez, P. (1994). A preliminary study of deaf educational policy. *Bilingual Research Journal*, 18(3-4), 85–113.
- Heift, T., & McFetridge, P. (1999). Exploiting the student model to emphasize language teaching in natural language processing. In M. B. Olsen (Ed.) *Proceedings of Computer-Mediated Language Assessment and Evaluation in Natural Language Processing, an ACL-IALL Symposium* (pp. 55–61). College Park, Maryland. Association for Computational Linguistics.
- Higgins, J. (1995). *Computers and English Language Learning*. Norwood, New Jersey: Ablex Publishing Corpo-

ration.

- Johnson, R. E., Liddell, S., & Erting, C. (1989). Unlocking the curriculum: Principles for achieving access in deaf education. *Gallaudet Research Institute Working Paper*, 89(3).
- Karttunen, L. (1983). KIMMO: A general morphological processor. *Texas Linguistic Forum*, 22, 163–186.
- Kelly, L. P. (1987). The influence of syntactic anomalies on the writing of a deaf college student. In A. Matsuhashi (Ed.) *Writing in Real Time: Modeling Production Processes*, Chapter 7, (pp. 161–196). Norwood, New Jersey: Ablex Publishing.
- Krashen, S. D. (1981). *Second Language Acquisition and Second Language Learning*. New York: Pergamon Press.
- Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. New York: Pergamon Press.
- Krashen, S. D. (1983). Newmark's "ignorance hypothesis" and current second language theory. In S. M. Gass & L. Selinker (Eds.) *Language Transfer in Language Learning*. Series on Issues in Second Language Research, Chapter 9 (pp. 135–153). Rowley, Massachusetts: Newbury House Publishers.
- Krashen, S. D., Butler, J., Birkbaum, R., & Robertson, J. (1978). Two studies in language acquisition and language learning. *ITL: Review of Applied Linguistics*, 39/40, 73–92.
- Larsen-Freeman, D. E. (1976). An explanation for the morpheme acquisition order of second language learners. *Language Learning*, 25(1), 125–135.
- Lesgold, A., Eggan, G., Katz, S., & Rao, G. (1992). Possibilities for assessment using computer-based apprenticeship environments. In J. W. Regian & V. J. Shute (Eds.) *Cognitive Approaches to Automated Instruction*, Chapter 3 (pp. 49–80). New Jersey: Lawrence Erlbaum Associates.
- Linton, F., Bell, B., & Bloom, C. (1996). The student model of the LEAP intelligent tutoring system. In *Proceedings of the Fifth International Conference on User Modeling* (pp. 83–90). Kailua-Kona, Hawaii. UM96, User Modeling, Inc.
- Loera, P. A., & Meichenbaum, D. (1993). The "potential" contributions of Cognitive Behavior Modification to literacy training for deaf students. *American Annals of the Deaf*, 138(2), 87–95.
- Loritz, D., & Zambrano, R. (1990). Using artificial intelligence to teach english to deaf people. Technical report, U.S. Department of Education, Office of Special Education and Rehabilitative Services, Technology, Educational Media, & Materials for the Handicapped Program. On Grant #H180P80020-89 to Georgetown University School of Languages and Linguistics, in consortium with Gallaudet University.
- Loritz, D., Parhizgar, A., & Zambrano, R. (1993). Diagnostic parsing in CALL. *CAELL Journal*, 1(4), 9–12.
- Mangelsdorf, K. (1989). Parallels between speaking and writing in second language acquisition. In D. M. Johnson & D. H. Roen (Eds.) *Richness in Writing: Empowering ESL Students*, Chapter 8 (pp. 134–145). New York: Longman.
- Matz, M. (1982). Towards a process model for high school algebra errors. In D. Sleeman & J. Brown (Eds.) *Intelligent Tutoring Systems*. Computers and People Series, Chapter 2 (pp. 25–50). Academic Press.
- McAnally, P. L., Rose, S., & Quigley, S. P. (1987). *Language Learning Practices with Deaf Children*. Boston: College-Hill Press.
- McCoy, K. F., Pennington, C. A., & Suri, L. Z. (1996). English error correction: A syntactic user model based on principled mal-rule scoring. In *Proceedings of the Fifth International Conference on User Modeling* (pp. 59–66). Kailua-Kona, Hawaii. UM96, User Modeling, Inc.
- Michaud, L. N. (2002). *Modeling User Interlanguage in a Second Language Tutoring System for Deaf Users of American Sign Language*. Ph.D. thesis, Dept. of Computer and Information Sciences, University of Delaware. Tech. Report #2002-08.
- Michaud, L. N., & 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*, Madison, Wisconsin.

- Michaud, L. N., & McCoy, K. F. (1999). Modeling user language proficiency in a writing tutor for deaf learners of English. In M. B. Olsen (Ed.) *Proceedings of Computer-Mediated Language Assessment and Evaluation in Natural Language Processing, an ACL-IALL Symposium* (pp. 47–54). College Park, Maryland. Association for Computational Linguistics.
- Michaud, L. N., & McCoy, K. F. (2000). Supporting intelligent tutoring in CALL by modeling the user's grammar. In *Proceedings of the Thirteenth International Florida Artificial Intelligence Research Society Conference (FLAIRS-2000)*, pages 50–54, Orlando, Florida. FLAIRS.
- Michaud, L. N., & McCoy, K. F. (2004). Empirical derivation of a sequence of user stereotypes. *User Modeling and User-Adapted Interaction*, 14(4), 317–350.
- Michaud, L. N., McCoy, K. F., & Pennington, C. A. (2000). An intelligent tutoring system for deaf learners of written English. In *Proceedings of the Fourth International ACM SIGCAPH Conference on Assistive Technologies (ASSETS 2000)*, Washington, D.C. SIGCAPH.
- Michaud, L. N., McCoy, K. F., & Stark, L. A. (2001). Modeling the acquisition of English: an intelligent CALL approach. In M. Bauer, P. J. Gmytrasiewicz, & J. Vassileva (Eds.) *Proceedings of the 8th International Conference on User Modeling*, Vol 2109 of *Lecture Notes in Artificial Intelligence* (pp. 14–23). Berlin: Springer.
- Michaud, L. N., McCoy, K. F., & Davis, R. Z. (2005). A model to disambiguate natural language parses on the basis of user language proficiency: Design and evaluation. *User Modeling and User-Adapted Interaction*, 15(1), 55–84. Special Issue on Language-Based Interaction.
- Moore, D. F. (1987). *Educating the Deaf: Psychology, Principles, & Practices*. 3rd edition. Boston: Houghton Mifflin Company.
- Opwis, K. (1993). The flexible use of multiple mental domain representations. In D. M. Towne, T. de Jong & H. Spada (Eds.) *Simulation-Based Experiential Learning* Vol 122 of *NATO ASI Series F: Computer and Systems Sciences* (pp. 77–89). New York: Springer-Verlag.
- Paris, C. L. (1988). Tailoring object descriptions to a user's level of expertise. *Computational Linguistics*, 14(3), 64–78.
- Paris, C. L. (1993). *User Modelling in Text Generation*. Frances Pinter.
- Paul, P. V. (1998). *Literacy and Deafness: The Development of Reading, Writing and Literate Thought*. Boston: Allyn and Bacon.
- Pennington, M. (1996). *The Computer and the Non-Native Writer: A Natural Partnership*. Written Language Series (Marcia Farr, series editor). Cresskill, New Jersey: Hampton Press.
- Ploetzner, R., Spada, H., Stumpf, M., & Opwis, K. (1990). Learning qualitative and quantitative reasoning in a microworld for elastic impacts. *European Journal of Psychology of Education*, 5(4), 501–516.
- Quigley, S., & Paul, P. (1984). ASL and ESL? *Topics in Early Childhood Special Education*, 3(4), 17–26.
- Quigley, S. P., & King, C. M. (1982). The language development of deaf children and youth. In S. Rosenberg (Ed.) *Handbook of Applied Psycholinguistics: Major Thrusts of Research and Theory*, Chapter 9 (pp. 429–475). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Quigley, S. P., Power, D. J., & Steinkamp, M. W. (1977). The language structure of deaf children. *The Volta Review*, 79(2), 73–84.
- Ragnemalm, E. L. (1996). Student diagnosis in practice; bridging a gap. *User Modeling and User-Adapted Instruction*, 5, 93–116.
- Rich, E. (1979). User modeling via stereotypes. *Cognitive Science*, 3, 329–354.
- Richardson, J. E. (1981). Computer assisted instruction for the hearing impaired. *The Volta Review*, 83, 328–335.
- Rose, S., & Waldron, M. (1984). Microcomputer use in programs for hearing-impaired children: A national survey. *American Annals of the Deaf*, 129, 338–342.
- Rueda, R. (1990). Assisted performance in writing instruction with learning-disabled students. In L. C. Moll (Ed.) *Vygotsky and Education: Instructional Implications and Applications of Sociocultural Psychology*,

- Chapter 17 (pp. 403–426). New York: Cambridge University Press.
- Russell, W. K., Quigley, S. P., & Power, D. J. (1976). *Linguistics and Deaf Children*. Alexander Graham Bell Association for the Deaf, Washington, D. C.
- Schneider, D., & 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*, Vol 2 (pp. 1198–1204). Université de Montréal, Montréal, Québec, Canada. COLING-ACL, Morgan Kaufmann Publishers.
- Schuster, E., & Burckett-Picker, J. (1996). Interlanguage errors becoming the Target Language through student modeling. In *Proceedings of the Fifth International Conference on User Modeling* (pp. 99–103). Kailua-Kona, Hawaii. UM96, User Modeling, Inc.
- Schwartz, B. D. (1998). On two hypotheses of "transfer" in L2A: Minimal Trees and Absolute L1 Influence. In S. Flynn, G. Martohardjono & W. O'Neil (Eds.) *The Generative Study of Second Language Acquisition*, Chapter 3 (pp. 35–59). Mahwah, NJ: Lawrence Erlbaum.
- Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12(1), 40–72.
- Self, J. A. (1988). Bypassing the intractable problem of student modelling. In *Proceedings of the 1st International Conference on Intelligent Tutoring Systems (ITS-88)* (pp. 18–24). Montréal, Québec, Canada.
- Selinker, L. (1971). The psychologically relevant data of second-language learning. In P. Pimsleur & T. Quinn (Eds.) *The Psychology of Second Language Learning: Papers from the Second International Congress of Applied Linguistics*, Chapter 4 (pp. 35–43). Cambridge: University Press.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10(3), 209–231.
- Selinker, L. (1992). *Rediscovering Interlanguage*. London and New York: Longman.
- Sleeman, D. (1982). Inferring (mal) rules from pupil's protocols. In *Proceedings of ECAI '82* (pp. 160–164). Orsay, France.
- Spada, H. (1993). How the role of cognitive modeling for computerized instruction is changing. In P. Brna, S. Ohlsson & H. Pain (Eds.) *Proceedings of AI-ED'93, World Conference on Artificial Intelligence in Education* (pp. 21–25). Edinburgh, Scotland. Association for the Advancement of Computer in Education (AACE). Invited talk.
- Stevens, R. (1980). Education in schools for deaf children. In C. Baker & R. Battison (Eds.) *Sign Language and the Deaf Community: Essays in Honor of William C. Stokoe*. National Association of the Deaf.
- Stewart, D. A. (2001). Pearls of wisdom: What Stokoe told us about teaching deaf children. *Sign Language Studies*, 1(4), 344–361.
- Strong, M. (1988). A bilingual approach to the education of young deaf children: ASL and English. In M. Strong (Ed.) *Language Learning and deafness* (pp. 113–129). Cambridge: Cambridge University Press.
- Suri, L. Z., & McCoy, K. F. (1993). A methodology for developing an error taxonomy for a computer assisted language learning tool for second language learners. Technical Report TR-93-16, Department of Computer and Information Sciences, University of Delaware.
- Swisher, M. V. (1989). The language-learning situation of deaf students. *TESOL Quarterly*, 23(2), 239–257.
- Tarone, E. (1982). Systematicity and attention in interlanguage. *Language Learning*, 32(1), 69–84.
- Vygotsky, L. S. (1986). *Thought and Language*. Cambridge, Massachusetts: MIT Press. Translation revised and edited by Alex Kozulin; originally published in 1934.
- Washburn, G. N. (1994). Working in the ZPD: Fossilized and nonfossilized non-native speakers. In J. P. Lantoff & G. Appel (Eds.) *Vygotskian Approaches to Second Language Research*, Second Language Learning, Chapter 4 (pp. 69–81). Norwood, New Jersey: Ablex Publishing Corporation.
- Watson, P. (1979). The utilization of the computer with the hearing impaired and the handicapped. *American Annals of the Deaf*, 124, 670–680.

- Wilbur, R. B. (1977). An explanation of deaf children's difficulty with certain syntactic structures of English. *The Volta Review*, 79(2), 85–92.
- Wilbur, R. B., Montanelli, D. S., & Quigley, S. P. (1976). Pronominalization in the language of deaf students. *Journal of Speech and Hearing Research*, 19(1), 120–140.
- Woolf, B., & McDonald, D. D. (1984). Building a computer tutor: Design issues. *IEEE Computer*, 17(9), 61–73.
- Woolf, B. P. (1984). *Context Dependent Planning in a Machine Tutor*. Ph.D. thesis, Department of Computer and Information Science, University of Massachusetts at Amherst. COINS Technical Report 84-21.