In this paper we briefly describe a project whose long term goal is the development of a “writing tutor” for deaf people who use American Sign Language (ASL). We envision a system that will take an essay written by a user, analyze that essay for errors, and then engage the user in a tutorial dialogue. We describe the overall design of the project, and then focus on one aspect which we feel is crucial for effectively tailoring corrective responses to the user: a model of the process of second language acquisition. We describe why such a model is crucial for effective tutoring, and briefly describe our model of the process.

Introduction

This paper briefly overviews a project whose long term goal is the development of a “writing tutor” for deaf people who use American Sign Language (ASL). We envision a system that will take an essay written by a user, analyze that essay for errors, and then engage the user in a tutorial dialogue designed to improve their overall literacy. The system is designed to be used by a particular student over an extended period of time. The tutoring that the system provides will be geared toward the specific student and his/her level of acquisition of written English.

Our preliminary work on this project has included an analysis of writing samples from deaf ASL natives. The analysis was intended to discover writing errors common to this population. While there were a variety of errors found in the sample analysis, our analysis supports the hypothesis that these writers are using the natural and beneficial strategy of building on their ASL knowledge when acquiring English. Our preliminary findings revealed that many of the errors (as much as 76% of those found in our initial sample analysis) could be attributed to Language Transfer (LT) from ASL to English, if LT is defined in the broad sense introduced in our previous work (see, for example, [Sur93],[MPS96]).

While this preliminary work suggests a method for designing a tutoring system for this task, care must be taken to design a flexible system that will allow various kinds of hypotheses to be tested. We wish to design a system that

1. is able to tailor its responses to the individual’s level of expertise in a way similar to the tailoring seen in human tutors
2. can be flexible (yet consistent) in its correction strategies so as to provide a testbed for assessing the efficacy of various kinds of strategies
3. has the capability of varying its generated text to the user’s level of acquisition so as to examine the efficacy of such tailoring

In other words, while our goal is to provide a usable and effective computer tutor, we wish to create a system that could serve as a testbed for evaluating hypotheses about effective ways to tutor a second language learner in general. Various works have been published concerning the acquisition of a second language and effective strategies for teaching a second language (see, [Bro94] for a nice overview of these issues, and [And93] for such issues in the context of deaf students). One goal of this work is to develop computational models able to carry out the general (often intuitive) strategies outlined in the literature. In doing so we will necessarily make these strategies more explicit and develop a tool for testing their efficacy in an environment that guarantees their consistent application.
In this paper we give a brief overview of our system design, and then concentrate on modeling the process of second language acquisition. This model will allow the system to take the user’s level of expertise (or level of acquisition of English) into account. We argue that the user’s level of acquisition of English is crucial for effectively diagnosing errors in the student’s writing, and for determining the set of errors to concentrate on in the tutorial correction. This model is also helpful in determining the basic content of the corrective response, and in generating the actual text of the response so that it provides examples of the constructions the user is currently attempting to acquire. Finally we describe our proposal for modeling the second language acquisition process.

Literacy Issues for People Who are Deaf

The problem of deaf literacy has been well-documented and affects every aspect of deaf students’ education. Since data on writing skills is difficult to obtain, we note that the reading comprehension level of deaf students is considerably lower than that of their hearing counterparts, “… with about half of the population of deaf 18-year-olds reading at or below a fourth grade level and only about 10% reading above the eighth grade level…” [Str88].

ASL is a visual-gestural language whose grammar is distinct and independent of the grammar of English or any other spoken language [Sto60], [BP78], [BC80], [HS83], [KB79], [BPB83]. The structure of ASL is radically different from that of English. It has been said to be closer to that of Chinese or the American Indian language Navaho. In addition to sign order rules (which are similar to word order rules of English), ASL syntax includes systematic modulations to signs as well as non-manual behavior (e.g., squinting, raising of eyebrows, body shifts, and shaking, nodding or tilting the head) for morphological and grammatical purposes [BC80], [Lid80], [Pad81], [KB79], [KG83], [Ing78], [Bak80]. The modality of ASL encourages simultaneous communication of information which is not possible with the completely sequential nature of written English.

One can imagine that these radical difference are difficult to overcome, but also consider that the deaf student has little to no understandable input in the language they are attempting to acquire. Note also that ASL has no accepted written form. Thus, just recognizing the correspondence between English orthography and “word concepts” is a major feat. One can see that learning written English is an enormous task.

It is anticipated that our system will be a tool for deaf people to use in improving these skills. For instance, this system would provide the user with feedback on his or her writing without involving a human teacher. Some students might prefer this mode of feedback since they would not risk feeling a “loss of face” as they might with a human tutor. The hope is that this will get the students to write more. In addition to the writing itself, we hope to gear the system’s responses to provide understandable input of constructions that the student is currently attempting to master. This aspect of our system design will perhaps prove most useful for this population in particular because of their lack of usable English input.

We should note that while there are “style checkers” and “grammar checkers” on the market, these programs do not satisfy the needs of the deaf. Educators of the deaf (and other people working with deaf people) report that such checkers frustrate deaf writers. Such tools do not catch many errors that are common in the writing of people who are deaf, and, at the same time, they flag many constructions that are not errors. We ran some of our writing samples from deaf subjects through a few grammar checkers, and we judged the results to be consistent with these reports.

Overall System Design

Figure 1 contains a block diagram of the overall system under development. The system, called ICICLE (Interactive Computer Identification and Correction of Language Errors), is designed to be a general purpose language learning tutor; however, we have focused on its application to deaf users of ASL acquiring written English, essentially as a second language.

![FIGURE 1. ICICLE Overall System Design](image-url)
The input/feedback cycle begins when the user enters a portion of text into the computer. The user’s text is first processed by the Error Analysis component which is responsible for tagging all errors found in a given input (some subset of these errors will be corrected in detail by the response generator). This component takes one sentence of input at a time. It first does a syntactic parse of the sentence using the grammar of English augmented with error production rules or mal-rules [Sle82], [WVJ78]. These mal-rules allow sentences containing errors to be parsed with the grammar, and enable the system to flag errors when they occur. The mal-rules themselves are derived from an error taxonomy (which resulted from our writing sample analysis) and other ASL information. The possible effects of ASL on the errors identified are captured in the Language Model. The effects from the acquisition of English as a Second Language are captured in the Acquisition Model. These two models affect a scoring mechanism which is used to identify a single parse (and set of errors) when multiple possibilities exist [MPS96].

If one of the syntactic mal-rules was used in the parse, the sentence contains (an) error(s) and the sentence with annotations from the error rule which was used will be passed to the Response Generator. If the error involved is a discourse-level error (e.g., an NP deletion), discourse information will also be passed to the Response Generator. If no syntactic error is found (or perhaps interleaved with the syntactic parsing), a number of semantic error rules will be run on the sentence. These rules will uncover errors such as the mixing of have and be (a common error in the samples we have analyzed). If such an error is found, annotations on the semantic rule that “fired” will be given to the Response Generator. Finally, the Error Analysis module is responsible for updating any discourse information tracked by the system (e.g., focus information). Once this information is recorded, the next sentence will be analyzed.

After all analyses are completed, the text, along with the error results and annotations from the error rules, will be passed to the Response Generator. The Generator component processes this information (along with data from the User Model and possibly the History Module) in order to decide which errors to correct in detail and how each should be corrected (including what language level should be used in any required instruction). The decision as to which errors to correct in detail will be based on information from the Assessment Module which contains a steps-in-language-learning model (described later) and other assessment factors. In addition, the User Model contains other user characteristics and evaluations that may also affect this decision.

The second decision that must be made in the Response Generator is which kind of correction strategy to use in actually generating the response. This decision is also affected by information stored in the User Model, Assessment Module, and History Module. The content of the response itself will be derived from the annotations on the errors that were passed in from the Error Analysis component; additional content for the responses may be provided by the ASL/English “Expert” (Language Model) or the Assessment Module depending on the tutoring strategy that is chosen. Finally, the responses will be displayed to the user who then has an opportunity to enter corrections to the text and have it re-checked. At the same time, information from the Response Generator will be used to update the recent and long-term “history” of the user. This knowledge can then be utilized to assess the user’s second-language ability and other user characteristics, and to evaluate the success (or failure) of the correction techniques employed thus far. Eventually, this process will develop a unique tutoring strategy that is capable of dynamically meeting the changing needs, preferences, and abilities of each user.

**Accounting for the L2 Acquisition Process**

In the remainder of the paper we focus on one aspect of our system: the Second Language (L2) Acquisition Model. In this section we argue why this model is crucial for both major system components. In the next section we describe our preliminary work in developing this model.

**Identifying Errors**

Before a system can effectively tutor a student who is learning a second language, the system must have the ability to classify an error the student has made to appropriately diagnose the problem the student is having. A problem that the system faces is that a given surface string may be syntactically interpreted in several ways. Recall that one way the system detects error is that it parses each sentence and determines whether or not a mal-rule has been used. Such a system must deal with the problem of multiple parses where some parses may contain mal-rules and others do not, some may contain different mal-rules than others, etc. Deciding between the multiple parses corresponds to deciding which errors (if any) the student had in the given sentence.
If we had a model of what the student had already acquired, what the student was currently acquiring, and what the student was most likely to acquire next, this could be used to select the most likely parse of the sentence in a principled fashion. A student is most likely to make errors in constructions s/he is currently acquiring [Vyg86]. Thus, given a set of parses, the one that is most likely to best describe the input is the one that contains mal-rules corresponding to errors in what the student is currently acquiring. Having a model of second language acquisition and a method for determining where a user is in this model, is crucial in detecting errors.

**Focusing the Correction**

Once errors have been detected, the system must determine:

1. which errors to focus on in the correction?
2. what basic content should be included in the corrective response to effectively explain the error?

For these tasks, a model of second language acquisition is also crucial. Our decisions concerning both of these aspects stem from work in second language acquisition and educational research. This work indicates that as a learner is mastering a subject, there is a certain subset of the material that is currently “within their grasp.” This has been called the Zone of Proximal Development (ZPD) by Vygotsky [Vyg86]. This general idea has been applied to assessment and writing instruction by [Rue90]. There is a similar principle outlined in [Kra81] with respect to second language acquisition. Intuitively the knowledge or concepts within the ZPD are “ready to be learned” by the learner. It is what s/he is currently in the process of acquiring.

According to the above literature, instruction and corrective feedback dealing with aspects within the ZPD may be beneficial. On the other hand, instruction or corrective feedback dealing with aspects outside of the ZPD will likely have little effect and may even be harmful to the learning process (in the sense that the user may become bored or confused by information that they are unable to comprehend or apply). Thus in our system we plan to concentrate on correcting errors that involve features that are at or slightly above the student’s level of acquisition.

Assuming that we have been able to decide which errors to focus on in the correction, the question still remains as to what information about the error should be included in the corrective response. The answer to this question depends on why the user made the error in the first place. Consider the following example found in one of our writing samples:

“My brother like to go...”

Most of us would agree that this sentence has a problem in subject-verb agreement. Because the subject is third-person singular, the present tense verb should be “likes.” In this example, the final “s” is missing. Notice that there are several reasons why a sentence such as this may be generated. Consider that it may have been generated because:

1. the student doesn’t know that such agreement exists in the language. That is, the student may be unaware that the form of the subject has anything to do with the form of the verb in such sentences.
2. the student is mistaken about the syntactic form the agreement takes. In this case, the student is aware that s/he needs to mark subject verb agreement, but does not know how to do so (or believes that s/he has already done so).
3. the noun should be in the plural form (but the student mistyped). In this case the student actually knows that subject-verb agreement exists in the language and knows the form that agreement should take, but simply mistyped the noun.
4. the verb should be in singular form (but the student mistyped). This situation is parallel to the one above.

Notice that very different kinds of content would be required to effectively correct the above error depending on the actual reason for making it. In the first case, some general tutoring should be given, explaining that agreement exists in the language, the circumstances in which the agreement needs to be marked, and the form the agreement should take. In case 2, only the form of the agreement needs to be explained. In cases 3 and 4, no tutoring should be given. The mistake should simply be pointed out to the student; any additional corrective feedback would likely annoy the student.

Knowing where the student is in acquiring the second language can help a system distinguish among the cases above. If subject-verb agreement is something that the student has not acquired and is not about to acquire, case 1 is most likely. If, on the other hand, it is currently in the ZPD (i.e., currently being acquired by the user), then case 2 is the more likely situation. Finally, either case 3 or 4 is likely if subject-verb agreement has already been acquired by the user.

**Tailoring the Generated Responses**

One of the few areas of general agreement among most second language acquisition (SLA) researchers is
that linguistic input at or near the user’s current second language proficiency is beneficial for the acquisition/learning process [Kra82], [Tar82], [Vyg86], [Hat83]. This principle may be expressed in different ways, but the idea is essentially the same.

For example, in Krashen’s “monitor” model of SLA [Kra81], [Kra82] he describes the learning process as a series of language level attainments. The level of the user’s current ability in the second language is designated i. According to Krashen, in order to facilitate quicker progress, the most helpful input will include features of the i+1 level.

For Tarone and others, this phenomena is referred to as foreigner talk. It describes the almost unconscious manner in which native speakers of a language automatically simplify their speech to accommodate second language learners. Input simplification is also a key premise in pidginization/creolization accounts of acquisition [Hat83]. For the most part, these types of simplification have been shown to be very helpful to the learner. However, Chaudron points out that there are a couple of “simplification” strategies that may not always be beneficial [Cha83].

Given this evidence, it is very important for the ICICLE system to have the ability to generate text at or near the user’s current second language proficiency. Again, having a model of what constructions the user is currently in the process of acquiring will make this possible.

Given such a model, one must consider that the process of generating a sentence to get across a particular meaning is a problem of choosing appropriate syntactic constructions that will get across that meaning in a syntactically correct fashion. During the process of generating, it may be the case that the generator has several choices of constructions for any one of which will effectively get the meaning across. Our basic idea is that whenever the generator is faced with such choices, it should choose a construction that would provide input for what the user is currently acquiring.

Consider a simple example of getting across the following information: the object CAT has the property of being colored BLACK and has the property of being HUNGRY. Note that (in addition to other sentences) either of the following sentences would get the content across:

- The black cat is hungry.
- The cat that is black is hungry.

These sentences result from different choices being made in the generator for getting across the BLACK property of the cat. In the first case the property is presented as an adjective; in the second, a relative clause is used. Currently, there is no real reason to choose one of these constructions over another. However, consider that if we knew that the user was currently attempting to acquire NPs containing adjectives, then the first choice would be preferable because it would provide an example of a construction the user was attempting to acquire. Because relative clauses are much more difficult to acquire than NPs containing adjectives, the second sentence would likely not enhance the user’s acquisition of the language. On the other hand, if the user was ready to acquire relative causes (and had already mastered NPs with adjectives), then the second rendition may prove more useful to the user in the context of a corrective response.

Thus, having a model of the acquisition process and using it to make choices left to the sentence generator might help the user acquire the second language faster. At the very least, this hypothesis should be tested in a controlled fashion. Of course, our sentence generator must have the ability to allow the model of second language acquisition to influence its decision making. Our preliminary investigations indicate that sentence generation based on a modified “FUF” system [Elh93] may fulfill our needs.

Modeling the L2 Acquisition Process

We are currently developing a computational model that captures the way that English is acquired (as a second language). In acquiring English as a second language, there is considerable linguistic evidence that the acquisition order of English features is relatively consistent and fixed regardless of the first language [Ing89], [DB74], [BMK74]. In addition to studies concentrating on second language acquisition, research in language assessment and educational grade expectations (e.g., [Ber88], [Lee74], [Cry82]) also suggests that languages are acquired in a relatively fixed order. This research outlines sets of syntactic constructions (language features) that students are generally expected to master by a certain point in time. This work can be interpreted as specifying groups of features that should be acquired at roughly the same time. For example, one would expect that the group of features that differ between a first and second grade reading level should be acquired together (i.e., between first and second grade).

We have attempted to account for the preceding results in a language assessment model called SLALOM (“Steps of Language Acquisition in a Layered Organization Model”). The basic idea of SLALOM is to divide the English language (the L2 in our case)
into a set of feature hierarchies (e.g., morphology, types of noun phrases, types of relative clauses). Within any single hierarchy, the features are ordered according to their “difficulty” of acquisition, reflecting their relative linguistic complexity. The ordering within feature hierarchies has been the subject of investigation in work such as [Ing89], [DB74], and [BMK74].

Within any single hierarchy, the features are ordered according to their “difficulty” of acquisition, reflecting their relative linguistic complexity. The ordering within feature hierarchies has been the subject of investigation in work such as [Ing89], [DB74], and [BMK74].

FIGURE 2. Language Complexity in SLALOM

Figure 2 contains an illustration of a piece of SLALOM. We have depicted parts of four hierarchies in the figure: morphological syntactic features, noun phrases, verb complements, and various relative clauses. Within each hierarchy, the intention is to capture an ordering on the feature acquisition. So, for example, the model reflects the fact that the +ing progressive form of verbs is generally acquired before the +s plural form of nouns, which is generally acquired before the +s form of possessives, etc.

Notice that there are also relationships among the hierarchies. This is intended to capture sets of features which are acquired at approximately the same time. These connections may be derived from work in language assessment and grade expectations such as found in [Ber88], [Lee74], and [Cry82]. So, for example, the figure indicates that while the +s plural ending is being acquired, so too are both proper and regular nouns, and one and two word sentences. During this time, we do not expect to see any relative clauses.

We anticipate that SLALOM, when fully developed, will initially outline the typical steps in acquiring English as a second language. This model will then be tailored to the needs of individual students via a series of “filters,” one for each user characteristic that might alter the initial generic model. For instance, it is possible that the specific features of the student’s L1 will affect the rate or order of acquisition of the L2. In particular, one would expect features shared in the L1 and L2 to be acquired more quickly than those which are not (due to positive language transfer). Another possible filter could reflect how much and what kind of formal instruction the student has had in written English. For example, if the student’s educational program stressed subject-verb agreement, this feature could have already been learned, even though others “before” it in the original SLA model may remain problematic.

In developing the language learning model and its filters, we plan to compare our initial model (derived from acquisition literature) with the writing samples that we have already collected. We also expect to seek input from English teachers of deaf students. Additionally, we hope to collect samples of teachers’ corrections and compare them to the models that will have been hypothesized.

Once the SLALOM model is complete we expect to rely on user modeling techniques to “place” the user within this model. The user’s placement will essentially determine the set of English features s/he is currently learning. The system will therefore concentrate its instruction on these features. In this way, the system instruction should be most beneficial to the user [Vyg86], [Rue90], [Kra81].

Current Implementation Status

Our some of our earlier work has reported on our analysis of writing samples that resulted in a taxonomy of errors expected from deaf people who use American Sign Language. On the basis of this taxonomy, we have developed a (limited) grammar of English which has been augmented with a number of mal-rules capturing some of the errors in this taxonomy. The implementation uses the bottom-up augmented context-free chart parser from [All95]. We are

2. We intend this figure as an illustration only. In particular our current research is focusing on identifying the precise hierarchies, orderings and syntactic features in the hierarchies, as well as relationships among the hierarchies.

3. Note that we have collected writing samples with some user information for the authors of each sample. While our analysis so far has been restricted to proficient ASL signers, samples from other deaf writers might help us determine what the ASL “influence” filter (for example) might look like since it would apply to one group of samples but not to another.
currently in the process of developing the scoring mechanism for the mal-rules and the model that captures second language acquisition. In addition, we are developing a Windows-based interface to the system to be used for testing purposes.\(^5\)

**Conclusion**

In this paper we have introduced a project concerned with the development of an intelligent computer-assisted language learning system that attempts to aid people who use American Sign Language in acquiring written English. We have described the overall system design, and have concentrated on how the system can tailor its correction to those constructions that the user is currently acquiring. Such tailoring requires a model of the process of acquiring a second language. We have shown how such a model might be captured computationally and have indicated how it can be used by the system.

**Acknowledgments**

This work has been supported by NSF Grant # IRI-9416916 and by a Rehabilitation Engineering Research Center Grant from the National Institute on Disability and Rehabilitation Research of the U.S. Department of Education (#H133E30010). Additional support has been provided by the Nemours Research Program.

**References**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Source</th>
</tr>
</thead>
</table>

---

4. We thank Xingong Chang and David Schneider for their work on the mal-rule grammar for the ASL writing project. Thanks also goes to Karen Hamilton for her implementation of the database used for our error analysis.

5. Thanks to Yu Zhang and Jeff Morriss for their work on the interface, and to Lisa Masterman for her work on the corrections currently in the system.


