A Tutor for Teaching English as a Second Language for Deaf Users of American Sign Language

Kathleen F. McCoy **Lisa N. Masterman**
CIS Department
University of Delaware
Newark, DE 19716
mccoy@cis.udel.edu and masterma@cis.udel.edu

Abstract

In this paper we introduce a computer-assisted writing tool for deaf users of American Sign Language (ASL). The novel aspect of this system (under development) is that it views the task faced by these writers as one of second language acquisition. We indicate how this affects the system design and the system's correction and explanation strategies, and present our methodology for modeling the second language acquisition process.

1 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 wish to address the particular difficulties faced by the deaf writer learning English and to create a system with the capabilities of accepting input via an essay written by a user (possibly several paragraphs in length), analyzing that essay for errors, and then engaging the user in tutorial dialogue aimed toward improving his/her overall literacy. The goal is a system designed to be used over an extended period of time, with the capacity to model the student's state of language proficiency and changes in that proficiency. The tutoring provided by the system would then be hand-tailored toward the individual user and his/her level of acquisition of written English.

Such a system must have several components. First, it must have the ability to analyze the input texts and determine what errors have occurred. It must then be able to select which of these errors to discuss with the learner, and in what order to discuss them. Finally, it must be able to generate appropriate corrective tutorial messages concerning the errors, keeping in mind both the goal of correcting this sample text and the larger objective of improving the overall literary of the student.

Concurrent with these explicit components, the system must be capable of constructing and updating a user model to be consulted in both the selection of errors to be corrected and the generation of corrective text. This user model would take into account a theory of second language acquisition which regards the process as a systematic revision of an internalized concept of the language to be acquired. Students would be placed within a model of climbing literacy, with language concepts rated as above, below, or within their current realm of acquisition, and the tutorial interaction tailored to this model.

In this paper, after motivating our specific application, we introduce the architecture of our eventual system and motivate its various components. After describing our current implementation status, we motivate the need for a model of second language acquisition. We finish with describing how we propose to model this process.

2 Literacy Issues for People Who are Deaf

The problem of deaf literacy has been well-documented and has far reaching effects on every aspect of deaf students' education. Though data on writing skills is difficult to obtain, we note that the reading comprehension level of deaf students is considerable 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..."(Strong, 1988)

Some Deaf people use American Sign Language (ASL). ASL is a visual-gestural language whose grammar is distinct and independent of the grammar of English or any other spoken language (Stokoe, Jr., 1960), (Baker and Padden, 1978), (Baker and

Cokely, 1980), (Hoffmeister and Shettle, 1983), (Klima and Bellugi, 1979), (Bellman, Poizner, and Bellugi, 1983). The structure of ASL is radically different from that of English, being much more similar to that of Chinese or the Native American 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 (Baker and Cokely, 1980), (Liddell, 1980), (Padden, 1981), (Klima and Bellugi, 1979), (Kegl and Gee, 1983), (Ingram, 1978), (Baker, 1980). The modality of ASL encourages simultaneous communication of information which is not possible with the completely sequential nature of written English.

In addition to radical differences in the structure of ASL and English, another obstacle to the ASL user acquiring English is the unique processing strategies he brings to the task (Anderson, 1993). The cognitive elements used to store signs in short-term memory are distinctively different from those used with a spoken/written language. Also, hearers of spoken language buffer the speech in order to process it together in words and phrases, but the buffer for visually observed data has a much quicker decay time than that of auditory or visual data, which leads to repetition and redundancy in signed languages that does not occur in the same manner elsewhere. Moreover, long, involved utterances of a manual language are parcelled into small parts that are recursively reinforced, referring back to previous details as each new piece of information is added, another charactertistic atypical of spoken language.

Adding to these difficulties is the fact that ASL has no accepted written form, eliminating the opportunity to establish literacy skills in a fluent native language and then transfer those skills to the new language being learned. Perhaps the worst difficulty for the deaf learner is that he has little to no understandable input in the language he is attempting to acquire. Thus, in addition to providing feedback on the student's writing, a tutoring system should be capable of offering sample understandable input using constructions that the student is currently attempting to master.

We anticipate that our system will address the unique needs of the deaf population in other ways as well. 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 explaining the difficulties faced by the deaf learner of English, we do not propose that ASL natives are fundamentally different from other learners of English as a Second Language; rather, we want to stress the view that English is, for ASL natives, a fundamentally different and challenging language, motivating the need to adopt a Second Langauge Acquisition strategy toward facilitating the learning process. There exist many obstacles to this process, some which are shared with other native language populations and some which are unique, such as the absence of the opportunity to have English input tailored to the personal level of acquisition and understanding of the learner. The system we propose attempts to address these needs as closely as possible within its own constraints (i.e., without the ability to converse with the learner in his native language).

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 individuals) report that such checkers, geared toward the errors of hearing writers, frustrate deaf students. Tailored toward the writing style of fluent, native English speakers, they 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.

3 Overview of System Design

Figure 1 contains a block diagram of the 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. While our current focus is on users of ASL, and thus some of the modules will be specific to the errors and difficulties of this learner population, our eventual goal is to have the language-specific aspects of the system to be excisable, allowing modules for different native languages to be inserted, so the system would eventually be usable for any learner of English as a Second Language.

The input/feedback cycle of ICICLE begins when the user enters a portion of text into the computer. The user's text is processed by the *Error Analysis* component which is responsible for tagging all errors. This component first performs a syntactic parse of a sentence using an English grammar augmented with error-production rules, or mal-rules

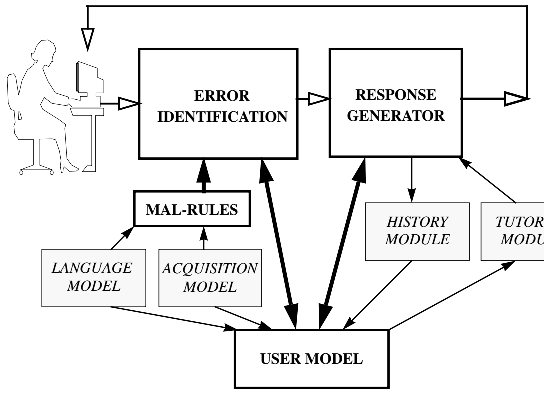

Figure 1: ICICLE Overall System Design

(Sleeman, 1982), (Weischedel, Voge, and James, 1978). 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 in conjunction with an analysis of how ASL knowledge might influence written English and other ASL information. The initial taxonomy was developed from an analysis of forty-eight Freshman and Sophomore writing evaluation samples from Gallaudet University (a liberal arts university for the deaf), seventeen writing evaluation samples from the National Technical Institue for the Deaf (NTID, a deaf school in Delaware), and five letters and essays written by ASL natives and collected through the Bicultural Center in Washington, DC. In total, the samples represent about 25,000 words. The errors were hand-counted and categorized, leading to the development of the mal-rules which represent them.

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* (described later in this paper). These two models affect a scoring mechanism which is used to identify a single parse (and set of errors) when multiple possibilities exist (McCoy, Pennington, and Suri, 1996).

The error identification phase must also look for semantic errors (e.g., mixing of `have` and `be`), and for discourse level errors (e.g., NP deletions). Some of these errors will be flagged after syntactic parsing using independent error rules. Finally, the Error Identification 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 generating any required instruction). The decision as to which errors to correct in detail will be most influenced by reasoning on the Acquisition Model.

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 and History Module. The content of the response itself will be derived from the annotations on the errors that were passed from the Error Analysis component; additional content for the responses may be provided by the ASL/English "Expert" (Language Model) and influenced by the Acquisition Model. 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.

4 Implementation

Our implementation to this point has concentrated most heavily on the analysis phase of processing. The user interacts with the system through a windows-based interface[1] through which text may either be entered directly or loaded from a file. Once the text is loaded, the user may ask that it be analyzed by the system.

The text is analyzed (one sentence at a time) by a bottom-up parser found in (Allen, 1995) using a grammar which has been augmented with mal-rules to capture errors uncovered in our analysis of writing samples.[2] The mal-rules are indexed with the errors that they realize.

The following is an example of a mal-rule from the grammar currently in implementation:

[1]We thank Robert Jeffrey Morriss for his work on the interface design and implementation.

[2]We thank Xingong Chang and David Schneider for their work on the grammar and Linda Suri for the writing sample analysis and development of the error taxonomy.

```
((s (inv -) (error-feature +) (wh ?w))
        -mv01.2>
        (np (case sub) (wh ?w) (agr p) )
        (head (vp (vform (? v pres past)) (agr s) (person 3)))
```

This rule would recognize an error at the sentence formation level, in subject-verb agreement–specifically, an error where the subject is plural but the verb form is third-person singular, such as "We does..." or "They has..." By tagging the sentence parse with the feature `(error-feature +)`, it is identified as containing an error, and the parse tree can be examined to discover the mal-rule (in this case, `mv01.2`) that was used in the parse.

After all of the sentences have been parsed in this way, the current system displays the text with colored highlighting over all error-containing sentences (different colors are used for different classes of error, again as identified from the mal-rules which were used). In addition, a color-coded menu appears which names the errors and associates them with the colors from the highlighted display. At this point the user may investigate the individual errors further. For example, s/he may click on a particular error name to get a (currently canned) explanation, or s/he may ask the system to mark the occurrences of a particular error only. In addition, the user may edit particular sentences, which results in an immediate new analysis of the text.

5 Accounting for the L2 Acquisition Process

There are several reasons why a model of second language acquisition is necessary.

5.1 Identifying Errors

It is common for our system to find multiple possible parses of an input string, where some parses may contain mal-rules and others do not, some may contain different mal-rules than others, etc. Deciding between these multiple parses corresponds to deciding which errors (if any) the student made in the given sentence. One area of our current work concerns progress toward making an *informed* choice about which parse tree best represents the student's input.

Our method is to develop a model of second language acquisition and use it for this task. For example, 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 (Vygotsky, 1986). 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 that realm of constructions (and that does not use constructions well beyond the student's current acquisition level).

5.2 Focusing the Correction

Once errors have been detected, the system must determine:

- which errors to focus on in the correction
- what basic content to include in the corrective response

Our model of second language acquisition is crucial for these tasks as well. Research in second language acquisition and education indicates that as a learner is mastering a subject, there is a certain subset of the material that is currently "within his grasp." This has been called the Zone of Proximal Development (ZPD) by Vygotsky (Vygotsky, 1986). This general idea has been applied to assessment and writing instruction by (Rueda, 1990), and second language acquisition by (Krashen, 1981). Intuitively the knowledge or concepts within the ZPD are "currently being acquired".

According to the above literature, instruction and corrective feedback dealing with aspects within the ZPD may be beneficial; 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, either boring or confusing the student with informtion he is unable to comprehend or apply. Thus the correction should focus on features at or slightly above the student's level of acquisition.

Once an error has been identified and chosen for a corrective response, the system must also decide on the content of that response. Here again, where the user is in the acquisition process (and thus, why he made the error) is crucial. Consider the following example found in one of our writing samples:

```
"My brother like to go..."
```

This sentence appears to most of us to have a problem in subject-verb agreement. Because the subject is third-person singular, the present tense verb should be "likes." Notice that there are several reasons why this error may be generated:

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 student intended the noun to be in plural form (but mistyped).
4. The student intended the verb to be in singular form (but mistyped).

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.

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. The student's placement in the model of acquisition can further direct our decisions regarding actions, because if this agreement is too far above the student's current level to be intellectually attainable at this time, we do not want to act on the error at all. If, on the other hand, it is currently *within* the ZPD) i.e., currently being acquired by the user), then case 2 is the most likely situation. Finally, either case 3 or 4 is likely if subject-verb agreement has already been acquired by the user.

5.3 Modeling the L2 Acquisition Process

We are currently developing a computational model that captures the way that English is acquired (as a second language) and gives us a framework upon which to project a student's "location" in that process. There is considerable linguistic evidence that the acquisition order of English features for second-language learners is relatively consistent and fixed regardless of the first language (Ingram, 1989), (Dulay and Burt, 1974), (Bailey, Madden, and Krashen, 1974). In addition to studies concentrating on second language acquisition, research in language assessment and educational grade expectations (e.g., (Berent, 1988), (Lee, 1974), (Crystal, 1982)) also suggests that language features 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 their study of the language. This work can be interpreted as specifying groups of features that should be acquired at roughly the same time.

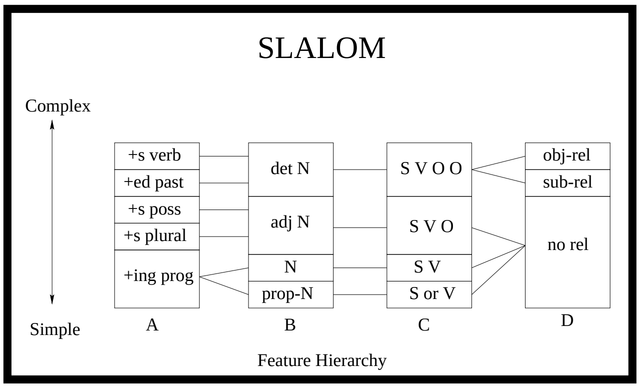

Figure 2: Language Complexity in SLALOM

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")[3]. 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, and 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 (Ingram, 1989), (Dulay and Burt, 1974), and (Bailey, Madden, and Krashen, 1974).

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 (Berent, 1988), (Lee, 1974), and (Crystal, 1982). The figure indicates that while the +s plural ending is being acquired, so too are both

[3]The initial work on SLALOM was done by Christopher A. Pennington.

proper and regular nouns, and one- and two-word sentences. While a learner is acquiring these features, we do not expect to see any relative clauses which are beyond that level of acquisition.

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 Native Language (L1) will affect the rate or order of acquisition of the Second Language (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 might reflect how various formal written-English instruction programs might alter the model, possibly stressing certain features normally acquired after others which remain unmastered.

We are developing the initial language learning model and its filters based on acquisition literature. We expect to further solidify the model using the writing samples that we have already collected. We are currently performing statistical analysis on our growing body of hand-corrected samples to see what error classes co-occur with statistical significance. We also expect to seek input from English teachers of deaf students, to see how they rank their students' abilities based on assignments they correct.

Once the SLALOM model is complete, we expect to rely on user modeling techniques to "place" the user within this model. This placement must be more sophisticated than simply looking at errors since some learners will avoid structures they do not know perfectly well in order to prevent error. Others will make heavy use of *prefabricated patterns*, such as the "tourist phrases" found in a travel book, whose use may precede a complete understanding of meaning or structure. Thus the placement algorithm must take into account both of these writing strategies.

6 Generating the Response

Aside from content, the generated response should have several other characteristics. In addition to providing examples of constructions the user is currently acquiring (as discussed earlier) the response should be organized so as to tie new knowledge into old knowledge thus facilitating *meaningful learning* as discussed by (Brown, 1994). When each new element is tied into already-learned data, and is presented so that pieces of new knowledge indtroduced together are related conceptually, the learning process gains a more significant meaning and new material is assimilate more quickly and entirely.

In addition, responses should encourage both *deductive* and *inductive* learning (where in the former, a standard practice for many foreign language classrooms, the student is introduced to the rule and is expected to use it to construct specific examples; in the latter the student is not directly told the rule, but is encouraged to generalize to the rule from specific correct examples). Classrooms benefit from both forms, but the deaf learner has limited to no exposure to correct forms, so responses that encourage inductive learning may be particularly useful. We postulate that this technique may be best achieved by providing positive examples from the student's own work. We have investigated the possibility of doing a search on the parse trees of correct sentences in the writing sample in order to find those that most closely fit a desired template, perhaps based on a sentence the learner has written incorrectly elsewhere.

The Response Generator should also take into account that feedback to a language learner occurs at two levels, affective and cognitive (Vigil and Oller, 1976). The *cognitive* level is that which concerns the content of the feedback, or the part which addresses the intellect of the learner and either enforces the assimilation of the concepts involved, or tells the learner to retry his attempt at communication. The *affective* level is less explicit, expressed through nonverbal cues and tone of voice, addressing a less conscious aspect of the learner. Negative feedback in this area should be avoided, as it may result in an abortion of his attempts to communicate. Even when the cognitive content of the response is indicating that an error occurred, the affective feedback should always encourage the learner.

7 Conclusions

It seems clear to us that the difficulties faced by deaf learners of written English require the development of such a tool as the one we envision. Direct, personalized interaction in a non-threatening (non-human) package, coupled with constructive input in the form of specific example utterances that address issues the student is currently learning, could go a long way toward bringing satisfactory English literacy within reach of the deaf population. Moreover, its general-purpose goals, stretching beyond this particular target audience of users, could make it a very useful tool for any language classroom.

8 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).

References

Allen, James. 1995. *Natural Language Understanding, Second Edition*. Benjamin/Cummings, CA.

Anderson, Jacqueline J. 1993. *Deaf Student Mis-Writing, Teacher Mis-Reading: English Education and the Deaf College Student*. Linstok Press, Burtonsville, MD.

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

Baker, C. 1980. Sentences in American Sign Language. In C. Baker and R. Battison, editors, *Sign Language and the Deaf Community*. National Association of the Deaf, Silver Spring, MD, pages 75–86.

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

Baker, C. and C. Padden. 1978. Focusing on the non-manual components of American Sign Language. In P. Siple, editor, *Understanding Language through Sign Language Research*. AP, New York, pages 27–58.

Bellman, K., H. Poizner, and U. Bellugi. 1983. Invariant characteristics of some morphological processes in American Sign Language. *Discourse Processes*, 6:199–223.

Berent, Gerald. 1988. An assessment of syntactic capabilities. In Michael Strong, editor, *Language Learning and Deafness*, Cambridge Applied Linguistic Series. Cambridge University Press, Cambridge / New York.

Brown, H. Douglas. 1994. *Principles of Language Learning and Teaching, Third Edition*. Prentice Hall Regents, Englewook Cliffs, NJ.

Crystal, David. 1982. *Profiling Linguistic Disability*. Edward Arnold, London.

Dulay, Heidi C. and Marina K. Burt. 1974. Natural sequences in child second language acquisition. *Language Learning*, 24:37–53.

Hoffmeister, R. J. and C. Shettle. 1983. Adaptations in communication made by deaf signers to different audience types. *discourse processes*, 6:259–274.

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

Ingram, R. M. 1978. Theme, rheme, topic and comment in the syntax of American Sign Language. *Sign Language Studies*, 20:193–218, Fall.

Kegl, J. and P. Gee. 1983. Narrative/story structure, pausing and American Sign Language. *Discourse Processes*, 6:243–258.

Klima, E. S. and U. Bellugi. 1979. *The Signs of Language*. Harvard University Press, Cambridge, MA.

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

Lee, Laura. 1974. *Developmental Sentence Analysis: A Grammatical Assessment Procedure for Speech and Language Clinicians*. Northwestern University Press, Evanston, IL.

Liddell, Scott K. 1980. *American Sign Language Syntax*. Mouton Publishers.

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

Padden, C. 1981. Some arguments for syntactic patterning in American Sign Language. *Sign Language Studies*, 32:239–259, Fall.

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

Sleeman, D. 1982. Inferring (mal) rules from pupil's protocols. In *Proceedings of ECAI-82*, pages 160–164, Orsay, France. ECAI-82.

Stokoe, Jr., W. C. 1960. Sign Language structure. *Studies in Linguistics occasional papers*, (8).

Strong, M. 1988. A bilingual approach to the education of young deaf children: ASL and English. In M. Strong, editor, *Language Learning and deafness*. Cambridge University Press, Cambridge, pages 113–129.

Vigil, Neddy A. and John W. Oller. 1976. Rule fossilization: A tentative model. *Language Learning*, 26:281–295.

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

Weischedel, Ralph M., Wilfried M. Voge, and Mark James. 1978. An artificial intelligence approach to language instruction. *Artificial Intelligence*, 10:225–240.