ANALYZING THE INTERLANGUAGE OF ASL NATIVES

by

Litza Stark

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Honors Bachelor of Arts in Computer and Information Sciences, with Distinction.

Spring 2001

ANALYZING THE INTERLANGUAGE OF ASL NATIVES

by

Litza A. Stark

Approved: ____
Dr. Kathleen McCoy, Ph.D.
First Thesis Reader

Approved: ____
Dr. William Idsardi, Ph.D.
Second Thesis Reader

Approved: ____
Dr. Nancy Jordan, Ph.D.
Third Thesis Reader

Approved: ____
Dr. Ann Ardis, Ph.D.
Director of the University Honors Program

ACKNOWLEDGMENTS

Thanks go to Jacy, my inspiration, partner, and guide for the duration of this work;
to my parents, for simply assuming that I could (and would) do anything I wanted to;
to Kathy, for steering me through this process and my college career;
and to Lisa, for providing consistent and persistent collaboration.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Deaf students learning English face a unique and daunting array of challenges because of the vast distance between signed ASL and written English. The ICICLE system would provide an automated tutor to address and remedy some of these difficulties. Central to the ICICLE system is the proposed user modeling component, which would parse input and provide tutorial feedback using (a) knowledge of the universal patterns of language acquisition, and (b) records of the individual user's past behavior. In order to develop a model of the syntax of Deaf students learning written English, we undertook this study. Using writing samples gathered from Deaf students that had been scored according to the Test of Written English, we tabulated individual syntactic errors on the sentence level and conducted MANOVA tests and factor analysis. We found differences in the occurrence of errors across proficiency levels based on the MANOVA test. From the factor analysis, we found classes of errors that were most characteristic of the proficiency levels, and which could be used to build a concrete model of the typical user's syntax.

Chapter 1

A TOOL TO AID DEAF LANGUAGE INSTRUCTION

1.1 Overview of the research

Along with others on the ICICLE team, I have conducted the following research into the acquisition process of Deaf students learning written English. This research was motivated by the needs of the ICICLE tool, an automated writing tutor designed to assist Deaf students learning English. In order for the system to represent a particular user and the path of his learning process, we need to know what specific stages the typical user passes through.

This research took two forms: we looked first at existing literature on Deaf students learning English to see if previous researchers had found patterns of acquisition. Then, having discovered that we still needed more information, we also undertook to find out for ourselves what those problems might be. In this research, we examined the errors that appear in the writing of a sampling of Deaf students at various stages of acqusition in order to characterize the intermediate syntax that they used.

1.2 Problems facing Deaf learners

Deaf learners of written English do not have exposure to the language in its natural, spoken state. Instead, they must learn the written form of the language in isolation. During their learning process, they cannot utilize the primary strategy of hearing learners: the correspondence between the written and the spoken word. Hearing children receive a staggering amount of linguistic input from the first moment they enter the world. In addition to the overwhelming quantity of input that they have at their disposal, the quality of this linguistic data may be specifically tailored to assist their language learning process, including special registers such as "motherese" and simplified speech. When it comes time for hearing children to learn written language, they no longer need to acquire all of their vocabulary or grammar, but simply have to transfer their auditory phonological knowledge to a written representation. In contrast, the only input Deaf learners have at their exposure is the written material, which may or may not be at an appropriate level for their stage of acquisition. They have none of the casual, ubiquitous exposure to language that hearing children do.

Deaf students with background in American Sign Language at least have the advantage of a solid, native-language background. However, even with this benefit, they still face unique challenges in learning English, resulting from problems inherent in the transfer from a visual and spatial mode to an auditory and written mode, as well as from their native language's linguistic distance from English. As a result, Deaf English education is in a dire situation, with "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). Students that already know ASL face the difficult task of transferring linguistic information from one perceptual mode to a completely new one. ASL is a visuo-spatial language in which linguistic features are carried not only by means of hand shape and movement, but also through facial expressions and body shifts. In making the transfer to written English, learners whose first language is ASL must become accustomed to a far more linear system in which simultaneous communication from multiple sources is not possible.

Besides the difficulties inherent in transferring between the two modes of speech, ASL is also structurally different from English, with syntactic constructions more akin to Chinese and Navajo (Michaud, Stark, & McCoy 2001). ASL and English differ in when and how certain features are marked (e.g., tense is only marked at the beginning of an utterance, and not on every verb), in the syntactic features present (e.g., there is no subject-verb agreement), and in the correspondence of lexical items (e.g., “other” and “another” are represented by the same sign) (Suri & McCoy 1993). Knowledge of ASL grammar does not in all cases facilitate understanding of English grammar in many basic syntactic categories.

1.3 A new tool: ICICLE

ICICLE (Interactive Computer Identification and Correction of Language Errors), a project led by Dr. Kathy McCoy, is a software system designed to be a general-purpose language tutor for acquiring English as a second language. Under its current implementation, however, it is tailored to attend to the particular needs of native speakers of ASL.

During an ICICLE session, as shown in Figure 1, the user begins by entering a piece of text and requesting that it be critiqued. The system uses both universal knowledge about the language acquisition process (the domain knowledge base) and specific information about this user’s past performance (the user model) to determine which errors the user has made in their writing. These errors are highlighted, and specific commentary and corrections are provided. In its intended implementation, the responses given to particular errors will also be tailored to the user's individual proficiency level. For instance, if the system knows that the user has used a given construction correctly in the past, it may simply need to remind the user of the correction. If, on the other hand, the user has never attempted the construction before, they might need a more in-depth description of the construction. The tutoring can continue indefinitely, with the user altering their text and the system reanalyzing the input and offering additional corrections. Information about a given session will be stored in the system so that it will know what to expect the next time it sees this user.

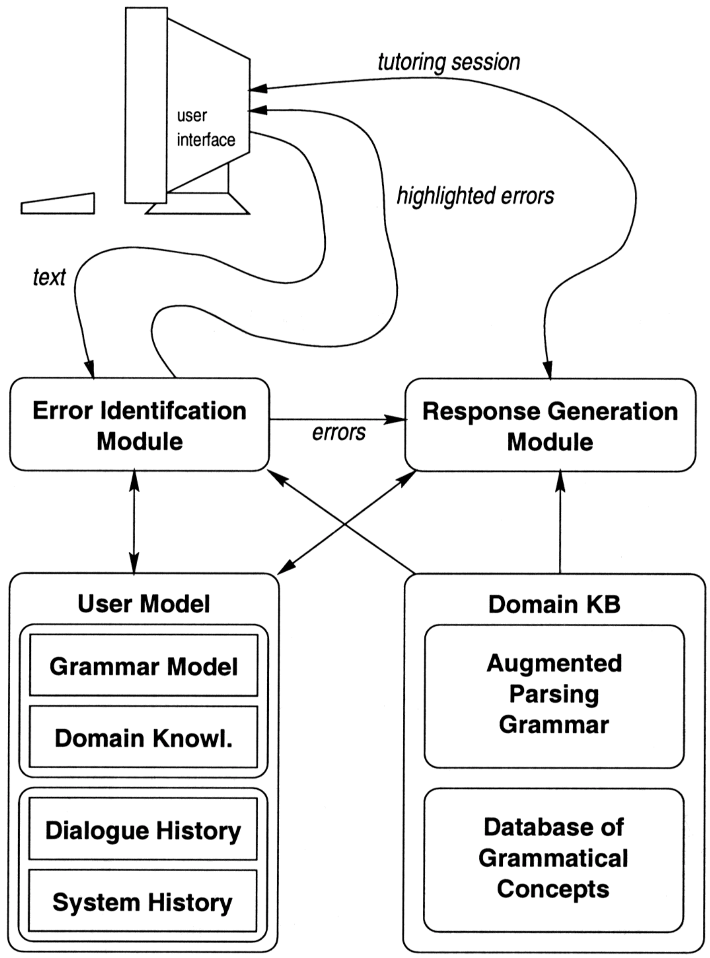

Figure 1 The ICICLE Architecture (from Michaud 1999)

The two primary tasks of the system, therefore, are to identify the errors in the inputted text, and then to generate an appropriate response to these errors. Both of these tasks rely on two underlying models: the domain knowledge base and the user model.

1.3.1 The domain knowledge base

In order to parse the user's input, the system must be able to parse both correct and incorrect inputs. It must, firstly, have a representation of correct English syntax. In addition, it needs a grammar of typical errors and their relation to correct constructions. Among these errors, some will be problems universal to learners of English as a second language, and some will be specifically related to the transfer of ASL grammar to English grammar. This knowledge base remains static across users, as it represents language universals that should be applicable to any learner.

1.3.2 The user model

Alongside the static knowledge base, the system must develop a model of each user. This user model should provide information for the dual functions of interpretation and feedback. In the case of ambiguous constructions in the input, a user model allows the system to choose the parse most likely to occur at the user's current stage of acquisition. And in designing feedback for the user, the user model can indicate both what corrections should be emphasized and the particular phrasing that should be used in the text of the feedback.

The user model will be a hierarchical representation of the user's linguistic development, referred to as SLALOM: Steps of Language Acquisition in a Layered Organization Model (McCoy et. al. 1996, Michaud 1999, Michaud et. al. 2001). Progress can be charted in specific areas, such as morphology, sentence structure, relative clause formation, or determiner use. Each one of these areas will constitute an individual hierarchy, in which elements are organized according to increasing complexity (See Figure 2 for a conjecture of how the SLALOM model might appear). For instance, within the morphology hierarchy, the "+ing" construction is learned before the “+s” plural construction, which in turn is learned before the “+ed” past tense construction. Individual vertical hierarchies will also be linked to one another horizontally; object noun relative clauses might be learned at the same stage as the SVO sentence construction, and the two hierarchies would thus correspond at this level. Much research has been done tracing the acquisition process of individual syntactic categories. However, not much has been done to link the hierarchies horizontally in order to describe which features are concurrently acquired. To develop a full SLALOM model, it is necessary to discover where the individual hierarchies are linked.

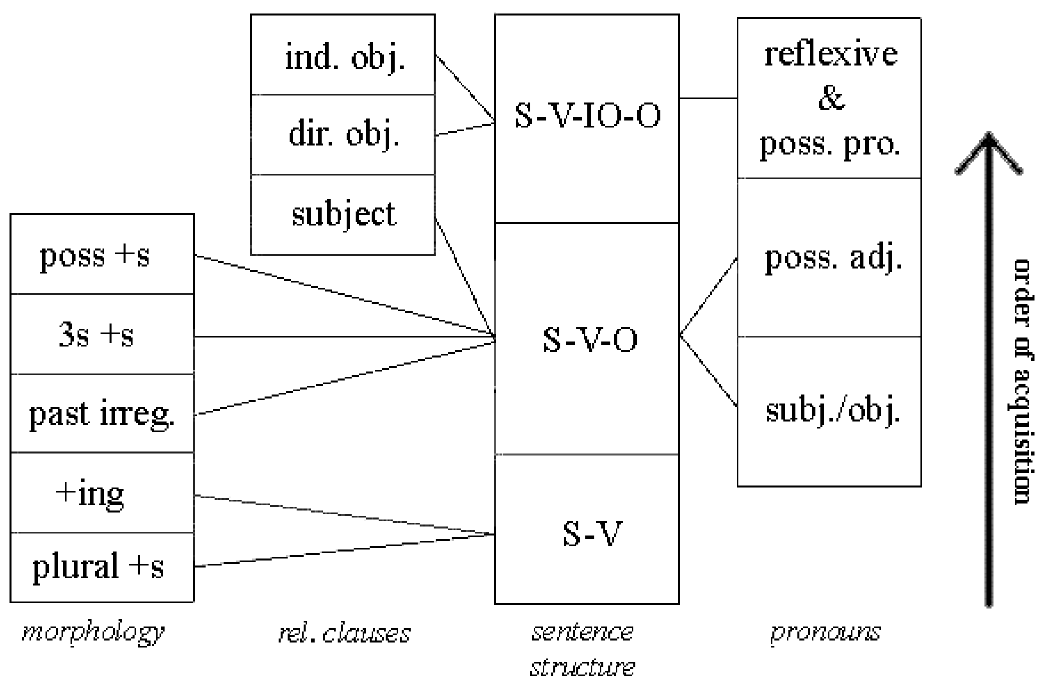

Figure 2 A hypothetical SLALOM hierarchy

Having such a model containing information about specific syntactic constructions would allow us to extrapolate predictions about one class of error based on observed errors of an entirely different sort. The system, based on SLALOM information, could designate a user as functioning at a particular level, and would expect their errors to come from the current level and (perhaps) the level above. It could assume that all constructions falling below the current level had already been learned. Having the information that SLALOM would provide could expedite the process of parsing a user's input and improve the specificity of the feedback offered to the user.

1.3.3 Linguistic assumptions behind the system architecture

In order for the ICICLE system to be feasible, its assumptions about the nature of language acquisition must have a valid theoretical basis. In the domain knowledge base, we must be able to formulate specific errors that will be likely to occur in the writing of ASL natives learning English. This module assumes that the errors made by second-language learners come in part from interference from their native language, and in part from the structure of their target language. The user model assumes that learners pass through fairly discrete and consistent stages in language acquisition, such that the errors in their writing will be predictable. In the following section I will provide evidence from existing research to suggest that we may safely operate under these assumptions.

Chapter 2

BACKGROUND IN INTERLANGUAGE AND ACQUISITION HIERARCHIES

2.1 Elements of interlanguage theory and the ZPD

Interlanguage theory, first proposed by Selinker in 1972, has become an increasingly accepted principle of second language acquisition. The theory describes the process of learning a second language as an evolutionary one, in which a learner passes through a sequence of intermediate grammars, or interlanguages, on the way to the target language (Selinker 1972). These interlanguages constitute the learner's working hypotheses concerning the grammar of the target language. The initial hypothesis is based largely on the grammar of the first language (L1). As the learner is exposed to more examples of the target language (L2), she revises her ideas about the L2 grammar. An ideal end state is reached when there is no mismatch between the language that she produces and the language that she is exposed to (Dulay & Burt 1974). Along the way, each interlanguage constitutes a grammar in its own right, rather than simply a scattershot attempt to mimic the L2. Each interlanguage consists of varying proportions of L1 grammar, L2 grammar, and hypothesized constructions that don't appear in either language.

During a learning process, there is always some material that has already been learned and some that is beyond the learner's grasp, but in between these two domains falls material that the learner is currently in the process of acquiring. Russian psychologist Lev Vygotsky called this material the Zone of Proximal Development

(ZPD), and suggested that this material constitutes a set of hypotheses that the learner is in the process of testing (Vygotsky 1986). She will make attempts to utilize her new concepts, and her usage will be full of errors that are not always consistent. It is this set of knowledge that an instructor should focus on; instruction on material beneath the ZPD is pointless, and instruction on that material beyond the ZPD will not take hold. The instructor should work with the student's natural learning process and help them conquer that material which is challenging them at a given moment. By using the SLALOM model, we attempt to pinpoint that material which is most likely to appear in the ZPD, and focus on instruction in that realm.

The interlanguage model of language acquisition provides the theoretical framework behind the SLALOM model, and Vygotsky's concept of the ZPD provides the instructional basis for ICICLE. Essentially, SLALOM's hierarchical levels represent different interlanguages, in which the earliest interlanguages contain the least complex structures. There is "an apparent predilection on the part of the learner for interlanguage structure which poses the fewest possible difficulties for mental processing" (Rutherford 1984). ICICLE seeks to take advantage of the natural hierarchical nature of language acquisition to facilitate its language instruction.

2.2 Sources of learners' errors

Traditional theories of second language acquisition hold that there are two sources for a learner's errors: language transfer and the influence of universal grammar. Thus, the *errors* arise out of interference from the first language and a learner's instinctual syntactical hypotheses. However, most of the *sequence* of the acquisition of any language, according to some researchers, is determined by the L2 itself. Studies on the acquisition of particular structures have found that the problems that learners of English as a second language have are frequently similar to those that young first-language learners of English have (Gass 1979). In addition, "regardless of first language background, children reconstruct English syntax in similar ways" (Dulay & Burt 1974). The native language causes variations on the standard sequence of learning a given language, but it is the target language that has the strongest influence on a learner's path of acquisition.

Universal grammar is another source of learners' difficulties. Some hierarchies of acquisition are not unique to a particular target language, but arise from universal cognitive limitations on the complexity of structures. Certain structures are inherently more complex than others, and thus take longer to be acquired cross-linguistically. Some research on individual syntactic structures (such as relative clauses) describes the particular hierarchy of difficulty for these structures. These hierarchies hold true not only cross-linguistically, in that more complex structures will never be present without also having the less complex ones, but also within an individual's learning process. Simple structures will be learned before more complex ones are attempted (Keenan & Comrie 1977, Keenan & Hawkins 1987).

2.3 Research in Deaf acquisition

The English acquisition process for Deaf students is not significantly different from that of a learner with any other first language (Swisher 1989). However, the influence of ASL and the lack of auditory information available to this population might offer some distinct difficulties. Berent, for instance, even concludes that there is a ceiling to Deaf students' capabilities brought about because of their processing limitations (Berent 1988). Whether or not there are such limitations, we are clearly dealing with a unique population of learners.

In describing the details of the SLALOM model, I had initially hoped to find existing research on the acquisition sequence of Deaf learners. A number of researchers have indeed focused on Deaf students learning English, but their research leaves some crucial gaps. Some studies reproduced acquisition research that had been done on first-language English learners and applied it to the learning process of Deaf students. These supported the claims of language universals in showing that Deaf learners undergo much the same processes as first-language English learners (Dulay & Burt 1973, Bailey et. al. 1974, Gass 1979). Some researchers conclude that, unlike learners with other first languages, Deaf learners have a ceiling to their possible language acquisition. They suggest that instructors should focus primarily on those structures that are most successfully acquired by Deaf learners, rather than wasting time on inherently difficult concepts (Bochner 1978, Berent 1988).

Quigley, et. al., have conducted the most comprehensive research concerning Deaf students' acquisition of English syntactic structures. His studies provide a fairly thorough picture of the timing of acquisition of individual structures. Interestingly, they focus exclusively on prelinguistically Deaf students, but don't investigate possible influences from ASL, nor do they specify what, if any, experience the subjects have had with ASL (though the majority of the subjects do come from residential schools for the Deaf). It is questionable, then, whether his research would be wholly reliable for making conclusions about the transfer from ASL to English (Quigley et. al. 1974, Quigley et. al. 1976, Wilbur et. al. 1976, Wilbur 1977, Quigley & King 1992).

Quigley tries to establish sequences of acquisition for a number of structures in the English of Deaf learners. He and his colleagues find rough sequences for questions, complement structures, verb tenses, and pronominalization. They find that the sequences closely match those found in first-language learners, although the acquisition is frequently delayed in Deaf learners. They do not seek to attribute errors to transfer from ASL, even when it might be appropriate. For instance, they note that the Deaf students have difficulty understanding the need to mark tense in both verbs of a conjoined sentence, but fail to mention that ASL does not require repeated tense marking within an utterance referring to a single topic.

Quigley provides a wealth of information about the acquisition of English by Deaf students with respect to individual constructions. His research supports the claim for a hierarchical sequence of second language acquisition with respect to Deaf learners. Many of his individual studies provide us with information about the sequence of acquisition within a given syntactic category. For the SLALOM model, however, we are seeking to link the acquisition processes of these constructions, and his data contain no temporal markers by which to compare the different processes. While his research is useful, we need to look elsewhere in order to describe the ways in which various structures interact during the acquisition process.

2.4 Implications for ICICLE

ICICLE and the SLALOM model rest heavily on the concepts delineated by interlanguage theory and the ideas of language transfer. The system will come equipped with expectations about rules derived from both English and ASL, in the form of the grammar and the rules defined for incorrect constructions. Based on the theory of language transfer outlined above, it would be unlikely for a user to use constructions that do not resemble either ASL or English grammar. Within the system, we seek to place a given user in a stage of acquisition that reflects an approximation of her current interlanguage. We try to emphasize those constructions that are being used erratically, as would be expected by Vygotsky's ZPD. Research in second language acquisition provides ample evidence to support the current design of ICICLE and the SLOLEM model.

Chapter 3

NEW ERROR ANALYSIS OF DEAF WRITING

3.1 Motivation for attempting error analysis

In order to create a useful SLALOM model, we need concrete data about the order of acquisition of syntactic structures by Deaf learners. Existing research, as outlined above, provides data concerning order of acquisition for independent structures, but we want to build a model that can express the interconnectedness of diverse syntactic structures. We need data that shows us what kinds of structures and errors tend to co-occur during the learning process, and whether they occur more frequently in writing from a learner at a particular stage of acquisition.

We undertook to conduct a new study to examine the acquisition of English syntax by Deaf students, paying close attention to the influence of ASL on their syntax. By conducting a broad study of the specific errors conducted by Deaf writing students, and comparing their fine-grained syntax with the overall proficiency of their writing, we hoped to establish a general outline for the SLOLEM model.

3.2 Intuitions and expectations

Upon first attempting this analysis, we had little idea how much to expect. We intended to count the errors found in the samples, to divide the samples according to a more global measure of proficiency, and then to look for distinctive patterns that would differentiate the various proficiency levels. These patterns would take the form of differences in the number or kinds of errors that occur at different levels. We then hoped that we could further examine these differences in order to find out what errors were particularly correlated with one another, and what errors were particularly distinctive to a given proficiency level.

3.3 Experimental design

3.3.1 Choosing the error measure

In examining the acquisition process and describing interlanguages, there are a number of indicators that might shed light on a writer's level of linguistic competence. The errors committed by a writer provide one of the most apparent indicators of the learning process, especially considering the assumptions behind the ZPD. Nevertheless, it is also important to examine those structures that are used correctly and those that are not attempted at all. Together, these three measures provide a complete portrait of a learner's interlanguage, and could potentially show us exactly where in the SLALOM hierarchy the learner falls. However, identifying those structures that are used successfully, or those structures that are not used at all, is a daunting task, and subject to much interpretation.

In an unconstrained set of writing, it may be impossible to determine those structures that are not attempted. However, it should be feasible to parse sentences either by hand or automatically in order to gauge which syntactic structures are being used successfully. At some point we hope to use the ICICLE parser to identify the correctly used structures, but now, while the parser is still under development, we are beginning by examining simply the errors committed in our corpus of data.

To enumerate the errors found in a given sample, we marked the errors that appeared sentence by sentence. We also grouped the samples according to their levels of proficiency. At this point, we could test to find any similarities in the error profiles among students at some proficiency level, or differences between the profiles of students at different proficiency levels. Assuming that the error profiles change along with the proficiency levels, this analysis should give us insight into which structures are being acquired at a given level.

3.3.2 The samples

All of our data come from an existing corpus of writing samples from Deaf students at Gallaudet University (GU), the National Technical Institute of the Deaf (NTID), the Pennsylvania School for the Deaf, and the Bicultural Center. Most of these samples (the bulk of them from GU and NTID) were written by entering college freshmen. These samples were collected completely confidentially; we do not know the writers' names or any identifying information. However, we screened out all samples except those most likely written by native ASL speakers[†], as we are particularly interested in examining the language transfer from ASL to English without interference from other signed communication systems. Our final corpus consists of 106 samples, with an average length of 17 sentences (228 words). (See Appendix A for a selection of writing samples.)

[†] In some situations, this was determined by a self-rating by the student; in other cases characteristics such as other Deaf family members or Deaf boarding school at an early age were used.

3.3.3 The error codes

We began with a set of 75 error codes, encompassing both syntactic and semantic problems. Each one of these codes could be identified at the sentence level, so that we could associate a list of codes with each sentence in a sample. In order to find the broadest results possible, the codes reflected many different syntactic structures, on the sentence level, on the phrase level, and on the word level.

3.3.4 The method

Each sentence was coded as a discrete unit, considering only the syntactic context of that sentence. Coders began their examination of a sentence by deciding which alterations were necessary to make the sentence grammatical by changing as few words in the sentence as possible. Once they had written a corrected form of the sentence, coders listed those error codes that described what led to the alterations, in the order in which they occurred. In the case of multiple interpretations for a given sentence, coders could list more than one correction and rate them as to their likelihood.

Table 1 Sample sentence and correction. Note that the error codes are listed in the order in which they occur in the sentence.

Original	People here are helpful and they make me to be self-confidence, mature.
Codes	**ev adjf mc**
Correction	People here are helpful and they make me self-confident and mature

Codes: **ev** = extra verb; **adjf** = adjective formation; **mc** = missing conjunction

Our final data consisted of the count of how many errors of each type occurred in each of the samples. Thus, we had a table of 106 entries (for each of the samples), each of which had 67 associated values (for each of the error codes).

3.4 Preliminary analysis

After the writing samples were completely coded using the set of 75 errors, we ran a preliminary set of statistical tests to look for patterns of errors. A single native-speaker judge classified the samples into four levels of proficiency based on subjective global assessment of the samples. The very worst samples verged on "word salad," showing very little grasp of English grammar; these were assigned a score of 1. Those that were slightly better—not quite unintelligible, with only rudimentary sentence structures—were assigned a score of 2. Samples with a score of 3 were intelligible, though full of grammatical and organizational problems. Sophisticated structures were occasionally attempted but not usually accurate. The best of the samples, which showed very few errors and were completely intelligible, were assigned a score of 4. After these ratings were given to the samples, the error occurrences within and across the informal levels were tabulated. Under this very casual analysis, we found an indication that there was a progression in the occurrences of errors across the proficiency levels (see Table 2).

Error	Proficiency			
	4	3	2	1
Incorrect determiner	■			
Punctuation error	■	■		
Sentence formation	■	■	■	
Verb tense formation	■	■	■	
Extra determiner	■	■	■	
No error	■	■	■	■
Semantic word choice	■	■	■	■
Missing determiner	■	■	■	■
Number error	■	■	■	■
Spelling	■	■	■	■
Incorrect preposition	■	■	■	■
Missing preposition	■	■	■	■
Verb tense/aspect		■	■	■
Missing subject		■		
Extra preposition		■		
Incorrect verb		■	■	
Unsure of intent		■	■	
Noun formation			■	
Missing verb			■	
Redundant		■		■
Missing auxiliary				■
Global sentence structure				■

Table 2 Preliminary analysis of common errors showed some trends towards differences among the four proficiency levels

Establishing that there was indeed some difference among the levels was not sufficient, however. It was also important to determine whether these errors were somehow related within the levels. Finding relations between errors at a given proficiency level will provide us with information about a particular horizontal cross-section of the SLALOM model, by showing us which syntactic categories are under acquisition simultaneously. To find those syntactic constructions which are in the ZPD at the same time, we looked for pairwise correlations of the occurrences of errors in the samples within each proficiency level. Eventually, if there were strong correlations between specific errors or error groups at a given proficiency level, the known presence of one error could help predict what other errors might be present, thus helping the user model to inference the full picture of the interlanguage on the basis of a few examples.

Indeed, there were a number of error pairs that were significantly correlated ($p<0.05$, with $0.40<r<0.90$) within each of the proficiency levels, and they tended not to be the same pairs of errors at each level. Interestingly, there were some errors that appeared to be the "hubs" of error clusters; these errors were correlated with as many as a dozen other error types within a given proficiency level. For example, incorrect verb errors at the fourth proficiency level (the highest level) were highly correlated with eleven other errors. And at the third proficiency level, be/have confusion errors correlated with ten other errors (See Appendix C for the correlations within the highest category, which is representative of how the data behaved in general across all of the levels). Statistically, this could frequently arise simply because the errors were widespread, occurring in most samples at a given proficiency level. However, these hub codes were not the most frequently occurring codes. The presence of significant correlations between codes, and the fact that these correlations were not identical from one proficiency level to the next, indicated that we were on the right track with our approach to the analysis. Encouraged by this preliminary research, we set out to hone our approach and strengthen our tests.

3.5 The final analysis

Reflecting on the preliminary analysis, we realized that there were a number of areas in which we could strengthen our design. The error codes were inconsistent and not always useful, the classification of the samples into broad categories had been too informal, and we had yet to establish that there was any strength to our coding process. In a second round of analysis, we sought to remedy these problems.

3.5.1 Refinement of the error codes

After the preliminary round of analysis, we realized that the existing set of error codes did not serve our purposes well. The errors described were ambiguous and overlapping, and in practice the various coders adhered to different meanings for the various codes. Additionally, the codes did not correspond well to the existing rules in the parser. If the errors were to be indicative of the acquisition of particular grammatical constructions captured in grammatical rules, this correspondence to the parser is an important point. Clearly, we needed to refine the set of codes and establish more rigorous guidelines for their application.

We examined each of the codes, checking where and how each had been applied in the previous analysis. We also decided whether it described an error that could be covered by an automatic syntactic parser, or whether it was context-dependent or semantically grounded and thus required human judgment to detect. We eliminated those errors that could only be detected with human judgment, and we explicitly defined the remainder. The remaining codes would reflect the parser's rules. In addition, explicit definitions of each error type were provided (with examples) so that the error codes could be systematically applied by different judges. We assembled a coding manual so that individual coders would not have to rely on their own assumptions or intuitions when applying error codes to a sample, and so that multiple coders could be trained and all coders would be working from a common set of guidelines.

The final set of 68 error codes, selected according to our preliminary analysis, fall into a variety of categories. Some are strictly syntactic errors (missing determiner, subject-verb agreement, question formation). A few reflect explicitly subjective problems with second-language learners, where they were absolutely necessary (ASL-ism, idioms). Still other error codes are less specific and perhaps more subjective (unsure how to parse, poor word choice, not a sentence). We hope that all of these classes of errors, taken together, could form a complete picture of language use that could eventually be applied in a syntactic parser. (See Appendix B for an explanation of each of the error codes.)

3.5.2 Coding and intercoder reliability

In order to establish that the set of codes and the coding process were objectively accurate, the two judges each coded a set of twenty samples for the purpose of comparison, using the new set of 68 codes and the guidelines set out in the coding manual. Such comparison presented a challenge, as the coders might disagree about the presence of an error, the location of an error relative to other errors, or the nature of the error itself. The strings of error codes associated with each sentence might not correspond exactly as to the errors that they described, as there might be gaps where one coder believed there was an ungrammatical construction and the other did not. Some examples of coder disagreement are shown in Tables 3 and 4.

Table 3 **Sample correction 1. Coder 2, because of a slightly different correction of the sentence, finds errors where Coder 1 did not. Where both coders find errors, however, they are in agreement.**

Original	But one thing worries me that most about NTID/RIT is financial problems.
Coder 1	**scf** **md** But one thing that worries me most about NTID/RIT is my financial problems.
Coder 2	**md** **scf** **md** **md** But the one thing that worries me the most about NTID/RIT is the financial problems.

Codes: **scf** = subordinate clause formation; **md** = missing determiner

Table 4 **Sample correction 2. The coders differ in the interpretation of a particular error, though both interpretations are valid.**

Original	I hope I could find some ways to solve.
Coder 1	**tc** **pl** **mo** I hoped I could find some way to solve them.
Coder 2	**sem** **pl** **mo** I wish I could find some way to solve them.

Codes: **tc** = tense context; **pl** = pluralization; **mo** = missing object

The evaluation of intercoder reliability, then, involved establishing a correspondence between the two coders' strings of error codes that accommodated gaps in the strings of errors listed by the two coders. This means that the different codes in Table 3 would be considered a 100% agreement, since the coders identified the errors in the same way when they actually found errors, and a 2/3 agreement in Table 4, since coders agreed in the identification of two out of three errors found. A string-matching algorithm that automatically inserted these gaps was used to compare the coders' performance (this algorithm was originally developed for use in comparing DNA sequences), and it was found that the Kappa value for the intercoder reliability was 0.78 (Michaud et. al. 2001). This is slightly below the standard acceptable reliability of 0.80, but we believe that this is an exceptional situation. There will always be inherent ambiguity involved in the assessment of grammaticality; in many instances, the coders delineated two very different, but equally correct sets of errors for a given sentence. Given these unavoidable differences of judgment, we believe that this rate of intercoder reliability is substantial enough to rely on in developing our user model. Since neither one of the coders can be considered to be "wrong," the interpretations that they give will not mislead the parser, and might even allow for a human-like ambiguity in the system itself. Having established sufficient reliability, the remainder of the samples were coded individually by one coder and the statistics were run on the results of these codes[†].

3.5.3 Classification of samples

The coding of the errors within sentences gave us a fine-grained assessment of the quality of the samples. As a more global evaluation of the proficiency of the writing, we turned to experts in the area of second-language instruction. Four English as a Second Language instructors at the English Language Institute of the University of Delaware examined the samples and rated them according to the standard six-point scale of the Test of Written English (TWE). Two instructors graded each sample; if their scores differed by more than one point, a third instructor also graded it. In any case, we assigned each sample the average of its assigned grade. Of our corpus, only one of the samples was scored a 1, and only four were assigned a 5. We combined these outliers with the next-closest categories for the purposes of our analysis. These scores served to classify the samples into roughly sequential stages in the second language acquisition process. (For a comparison of the different proficiency levels, see Appendix A.)

[†] For the twenty that were coded by both judges, ten from each coder were included in the final set of data.

3.5.4 Analysis

3.5.4.1 Differences across proficiency levels

To establish that there were indeed differences in the occurrence of errors across the levels of proficiency, MANOVAs were performed using the TWE scores as the independent variable and the normalized† error occurrences as the dependent variables. The MANOVA test seeks to find the effects of a single independent variable on multiple dependent variables. In this case, we have 68 dependent variables, in the form of the normalized counts for each of the error types, and the single independent variable of the overall proficiency score assigned to a sample. The MANOVA test should tell us whether the error types have a significant effect on the score of a sample, and how large those effects are. Thus, it is a way of testing whether the counts of error codes can indeed capture concrete differences between samples at various proficiency levels.

The MANOVA examined the occurrence of each error code at each of the TWE score levels. Of course, not all errors showed significant differences between every proficiency level, but results strongly suggested that many errors occurred at different rates within the various proficiency levels. Using the number of error occurrences normalized by the number of words in a sample, we found that there were significant differences between proficiency levels for 24 out of the 68 error types ($p<0.05$). Table 5 shows the errors that show significantly different rates of occurrence from one proficiency level to the next. These results encouraged us that we

† For our analyses, we normalized the number of occurrences of errors by the number of words in the sample, so that longer samples would not be disproportionately weighted.

might be able to find even more detailed information about the patterns of errors that characterize a given proficiency level. For a fuller discussion of the MANOVA results, see Michaud & McCoy 2001.

Table 5 Di fferences between errors between proficiency levels. Bold denotes p<0.01 and normal denotes p<0.05. Values within cells are the differences in error counts between levels.

ID	Error	2 vs 3	2 vs 4	3 vs 4
1	none		**-4.37**	**-3.74**
3	unsure how to parse	**0.81**	**1.46**	0.65
4	random error		**1.47**	**1.01**
6	incorrect determiner			0.53
7	missing determiner			2.62
11	extra pronoun		**1.04**	**0.89**
15	adj/adv confusion			**0.42**
16	adjective formation		**0.8**	**0.71**
21	pluralization problem	1.52		
22	pronoun error		**0.93**	**0.78**
32	infinitive error		**1.07**	**0.8**
34	activity		**1.07**	**0.97**
35	tense context error		**4.7**	**3.75**
39	missing verb		**1.28**	**1.36**
46	missing object		**0.61**	**0.62**
52	incorrect relative pronoun			**0.32**
60	missing conjunction		**1.58**	**1.03**
67	idiom		0.67	0.51
68	ASLism		**0.11**	**0.09**

3.5.4.2 Patterns within proficiency levels

Before discussing the results of the factor analysis that we performed on the data, I will briefly explain the procedure and its implications. Factor analysis seeks to simplify large sets of data by reducing the number of dimensions being measured, producing instead a few measures that (ideally) capture underlying causes of the measured variables.

Consider an analysis of writing samples in which three measures are taken: the number of words in a sample, the average sentence length, and the number of distinct words used (the vocabulary). We might expect these measures to be rough indicators of the sophistication of an essay, and therefore that they would be highly correlated. We could represent the relations among the measures by plotting each individual essay as a point on three-dimensional coordinate axes. On this graph, each axis would represent one of our measures. The points, since the measures are highly correlated, would cluster in roughly a football shape along a three-dimensional diagonal. This line through the football's long axis captures the essence of the figure, expressing in shorthand the general trend of the data. If the three measures are perfectly correlated with one another, then this line will capture every data point. Since they are not perfectly correlated, many of the data points will lie farther from the axis, and we will need some additional description to capture the behavior of the data. We can draw another line perpendicular to the first that captures somewhat more of the variation in the data, though not quite as much as the first line that we drew. These two lines represent in shorthand the behavior of the data as a whole (Gould 1981).

By performing factor analysis on data containing a set of *n* variables, we are essentially plotting our data as a football in *n*-dimensional space, and then drawing axes to describe the shape of the data in a more simplified way. The first axis that we draw captures most of the behavior of the data, and subsequent axes account for decreasing amounts of variance, or distance from perfect correlation. A small number of axes constitute a sufficiently detailed representation of the data, while simplifying the number of variables that we need to use.

We can see what is happening when we reduce the data into individual factors, but what are these factors and how do we arrive at them? To conceptualize what an individual factor consists of, consider the measure of each error type to be represented by a vector of unit length, with all vectors radiating from a common point. If two measures are highly correlated, then the angle between them approaches 0º. If they are completely independent, then their vectors lie at right angles to one another. When we attempt to reduce the individual measures into a factor, we are consolidating a group of measures that are highly correlated with one another, and that therefore display similar patterns of behavior. The first component that we choose is the one that has maximal resolving power. This resolving power is calculated by projecting the vector representing the measure onto the factor (see Fig. 3). The ratio of the length of this projection to the length of the original vector measures the percentage of the vector's information resolved by the factor. If all of the measures are highly correlated with one another, then the first component (the principal component) resolves most of the information in the individual measures. If the measures are not so as highly correlated, then additional components will capture the remaining information (Gould 1981).

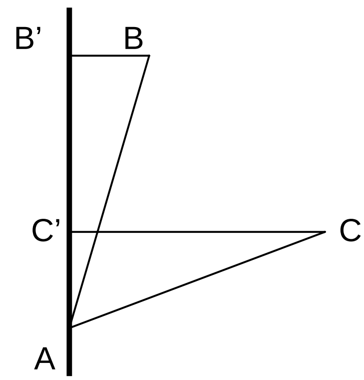

Figure 3 **Projection of vectors onto an axis. "The amount of information resolved by the axis is the ratio of the projected length on the axis to the true length of the vector…Vector AB lies close to the axis and the ratio of the projection AB' to the vector itself, AB, is high. Vector AC lies far from the axis and the ratio of its projected length AC' to the vector itself, AC, is low" (Gould 1981).**

When we have a factor that accounts for a significant amount of the behavior of our data overall, this factor may represent a better abstract underlying source for the behavior of the data than any one of the individual measures. In the simplified example I used above, the principal factor resolving the length of an essay, the average sentence length, and the vocabulary size might represent the underlying sophistication of the essay. In performing factor analysis on the data from our samples, one proficiency level at a time, we are seeking to find factors composed of groups of errors that are highly correlated with one another. These errors, because they are able to describe much of the behavior of the data as a whole, may indicate the abstract underlying character of writing at a given stage of acquisition. This will be particularly evident if we find that the error types that constitute our factors are consistent with earlier research on language acquisition.

We performed a factor analysis both with the full collection of error codes and with a reduced set of 21 codes yielded by grouping the original 68 codes into general syntactic categories (see Appendix B for a list of the 21 sets). Using the full set of 68 error codes, the analysis produced nearly as many factors as there were samples in the set, simply because there was not enough data to calculate strong factors. Therefore, the results of the reduced factor analysis are probably stronger. See Table 5 for a breakdown of the factors computed.

Table 6 Results of factor analysis using reduced set of 21 errors, using cutoff point of 0.40 for principal components

All Samples (N=106)

Factor 1	Factor 2	Factor 3	Factor 4
1 - SS: SVO 4 - SS: conjunctions 6 - NP: det N 7 - NP: prep NP 8 - NP: Adj N 9 - Morph: N case 11 - Morph: modifiers 14 - Morph: V +ing 18 - VP: inf V	10 - Morph: N +’s 13 - Morph: V +s 14 - Morph: V +ing 17 - VP: aux + V	3 - SS: S V Adv 19 - Dummy subjects	5 - SS: questions 19 - Dummy subjects
Factor 5	**Factor 6**	**Factor 7**	**Factor 8**
16 - VP: Negation	15 - Morph: V +ed 20 - Sub. clauses	15 - Morph: V +ed	12 - Morph: V

Low Samples (N=42)

Factor 1	Factor 2	Factor 3	Factor 4
1 - SS: SVO 4 - SS: conjunctions 6 - NP: det N 7 - NP: prep NP 8 - NP: Adj N 11	1 - SS: SVO 12 - Morph: V 14 - Morph: V +ing 17 - VP: aux + V 18 - VP: inf V	10 - Morph: N +’s 13 - Morph: V +s 19 - Dummy subjects 21 - Comparisons	6 - NP: det N 7 - NP: prep NP 10 - Morph: N +’s
Factor 5	**Factor 6**	**Factor 7**	
13 - Morph: V +s 17 - VP: aux + V	3 - SS: S V Adv 9 - Morph: N case	20 - Sub. clauses 21 - Comparisons	

Medium Samples (N=38)

Factor 1	Factor 2	Factor 3	Factor 4
1 - SS: SVO 3 - SS: S V Adv 7 - NP: prep NP 9 - Morph: N case 10 - Morph: N +’s 13 - Morph: V +s 14 - Morph: V +ing 17 - VP: aux + V 18 - VP: inf V	9 - Morph: N case 15 - Morph: V +ed 19 - Dummy subjects	1 - SS: SVO 16 - VP: Negation 20 - Sub. clauses 21 - Comparisons	8 - NP: Adj N 12 - Morph: V 18 - VP: inf V
Factor 5	**Factor 6**	**Factor 7**	
5 - SS: questions 21 - Comparisons	4 - SS: conjunctions 16 - VP: Negation 20 - Sub. clauses	2 - SS: S V PP	

High Samples (N=26)

Factor 1	**Factor 2**	**Factor 3**	**Factor 4**
1 - SS: SVO 3 - SS: S V Adv 4 - SS: conjunctions 6 - NP: det N 7 - NP: prep NP 11 - Morph: modifiers 14 - Morph: V +ing 20 - Sub. clauses 21 - Comparisons	7 - NP: prep NP 9 - Morph: N case 10 - Morph: N +'s 14 - Morph: V +ing 15 - Morph: V +ed 17 - VP: aux + V	1 - SS: SVO 12 - Morph: V	17 - VP: aux + V
Factor 5	**Factor 6**	**Factor 7**	
6 - NP: det N 9 - Morph: N case 18 - VP: inf V	13 - Morph: V +s 21 – Comparisons	3 - SS: S V Adv	

We performed factor analysis of the entire set of samples to provide a basis of comparison, permitting us to rule out factors that might be more reflective of the samples as a whole than of any individual proficiency level. If the principal factor of the entire set of samples has a high degree of overlap with the principal factor of the subgroups, then we can conclude that this factor is accounting for characteristics of the writing in general, rather than of any proficiency level specifically. Within the first factor, there was indeed a great deal of overlap among the various levels—error codes 1, 4, 6, 7, 11, and 14 appear in at least three of the four levels. These types of errors, then, are particularly characteristic of the samples in general, and account for much of the variation among the samples, but they cannot be used as a set to characterize any of the individual proficiency levels.

The second factor, which also accounts for a great deal of the variation in the data, yields more information. There is not such a high degree of overlap between the various levels for this factor, so we can conclude that these factors are capturing patterns within the data that are distinct to the individual proficiency level. Using the samples that appear in this factor, we can lay out a set of some of the errors that characterize each successive ELI proficiency level.

Table 7 SLALOM model based on factor analysis (using error categories from second factor)

	Sentence Structure	***Morphology***	***Verb Phrases***	***Other***
ELI 4		N case N +'s V +ing V +ed	aux + V	prep NP
ELI 3		N case V +ed		dummy subj
ELI 2	SVO	V V +ing	aux +V inf V	

Table 6 contains the error categories included in the second factor for each of the proficiency levels. They have been divided into broad categories of linguistic features in order to express the progression from simpler structures to more complex ones, just as we aim to do in the SLALOM model. In the morphology category particularly, we see a trend towards the progression from more simple morphological constructions (V at the ELI rating of 2, for instance) to more complex ones (N +'s at the ELI rating of 4). These are the errors that are most pronounced at the given proficiency level[†], and, based on our assumptions, these are the constructions that are currently in the ZPD.

There is not enough information based on this test to build a robust SLALOM model that can account for the entirety of the interlanguage grammars. In addition, there are some inconsistencies, such as the presence of some error types at levels 2 and 4 that do not appear in level 3. Based on our assumptions, the same errors

[†] It should be noted that these errors are not necessarily the most frequently occurring at these proficiency levels; they are simply the groups of errors that are most characteristic of the proficiency levels.

may appear at successive levels, but that once they disappear they should not appear again. This may not be in violation of our expectations, however. Our combination of many errors into larger categories may be masking the true structures being acquired. For example, we see “ing” construction problems at the first level and the third level, but disappears in the second level. Its appearance at the earlier level may reflect the acquisition of simple verb construction (“He was going to the store”) while its appearance at the later stage may be for the more sophisticated adjectival use (“The man going to the store grinned”). Thus, one place where more research can be done is in looking more closely at places such as this in an attempt to tease out the actual constructions being acquired.

Nevertheless, the results of the factor analysis are encouraging in that they indicate the presence of distinct syntactic patterns at the different proficiency levels, and that these patterns can be captured by such fine-grained analysis of the errors that occur within samples of writing.

Chapter 4

CONCLUSIONS AND FUTURE WORK

4.1 Conclusions

The results of the MANOVA indicate that we can indeed measure the differences between levels of language acquisition by examining the individual errors that appear in writing. This offers encouragement for further attempts to describe the acquisition process by looking at the interactions of diverse syntactic structures.

The factor analysis, while showing that the writing has at its core very similar patterns of syntax, does indicate that there are errors that are particularly indicative of given proficiency levels. These findings provide support for our experimental method, and also provide us with new information for the development of the rest of the ICICLE system. The factors produced by the analysis show us what groups of errors are most likely to co-occur at a given proficiency level. Knowing some of the characteristic errors at a given level (the secondary factors for each proficiency level) may help in disambiguating sentences when the user's level is known. The errors included in the primary factor might also be helpful for disambiguation when the user's level is not known at all. These results indicate that such factor analysis may be utilized to capture a more robust model of a user's interlanguage syntax.

4.2 Limitations of this approach

As noted earlier, there are a number of ways in which this study is incomplete. While focusing exclusively on errors provides a streamlined way of analyzing interlanguage syntax, a complete portrait of a grammar can only be created from information not only about unsuccessful constructions but also about successful ones. For the factor analysis particularly, it will be helpful to include both data on the errors committed and data on the structures successfully executed. In the near future, we hope to incorporate such data by using the parser to determine what structures actually appear in the samples.

Our corpus of samples represents a diverse set of students, but for the most part each of the samples comes from a different individual. To get a more realistic perspective of the development of interlanguages within an individual, it will be necessary to do within-subject longitudinal studies of the writing of Deaf students. In general, more data will strengthen the information that we have already been able to collect.

4.3 Towards a full user model

The results of this study will be useful in developing the user model for the ICICLE system. From the MANOVA tests, we know that there are measurable differences between the proficiency levels of our potential users, and the factor analysis has started to delineate what these differences are. The factors yielded by this and future factor analysis will provide us with some characterization of the syntax of a user at specific proficiency levels. The influence of the hierarchical user model may be invisible to the individual user, but its influence within the system may be far-reaching.

Appendix A

THE WRITING SAMPLES

A.1 Sample with rating 2

I arrived on thursday, I lost my Parent's way but I evade other Parent car.

I waited in line for registration paper.

They gave a message phone from Parent call.

I was shocked, felt thank ask me so I stayed while few minutes, Parent came here in final.

Parent told me, They was anxious to lost his son.

I felt a comfortable to be here.

That's why I was quafity in sign also like in Pidgin Sign English.

I was glad to carried thing and lugguage think put in room.

I meet new student's communication in the dormitory.

My Parent stayed here for 3 days with me, Parent was sign and observe to see where they are on place.

I was interesting to walk out on tunnel of basement.

I noticed there were different rooms.

Tunnel (basement) were nice places, store, cedar, Flaming.

I was remember in my old friends but I can't see for 2 years, I was surprised.

That's who was interpreters, It was friend, Later I was walking around in LBJ Building, I found other old friend last 2 months.

My old friend was interpreter.

I hope so Next Month There was some old friend will come here.

A.2 Sample with rating 3

What do you like about NTID and the people here?

I am glad that I chose the right college for me to go this year because

I heard good things about NTID and a major that what I want to take.

I like to go techolony college.

NTID is really a nice place and good college.

I feel that it is a good experience for me to go NTID.

I enjoy learning alot of things and experiences at NTID.

I have to learn how to be an independent myself at dorm and NTID.

I have to learn alot of responisibles at NTID.

I want to have a good life at NTID.

NTID helps students learn alot different experiences and solve their problems.

NTID has tutors and counselors to help the students with their classes, problems with their friends or lives, or anything if the students need them help or explain.

I met alot of students at NTID, SVP 85 and it was enjoyable.

The students are friendly and talkative.

We help with each other with problems or anything.

It is experience for me to meet many different people at SVP 85.

It is great to know and talk with new friends.

It will be more people at NTID next fall.

A.3 Sample with rating 4

Baseball and ice hockey have very little in common and have lots of differences.

Both sports are very popular and exciting to watch.

Both sports have rules to follow.

Baseball has twelve innings and can have more innings until the tie is broken.

Ice hockey has three periods and has a limited overtime.

Sometimes the game stays tied up.

Ice hockey has a penalty box for any player who fouls.

Baseball only has three strikes and the player is out.

The teams switch to either defense or offense when there are three outs.

Baseball and ice hockey have equipment that they use while playing.

In baseball, players use baseballs, bats, bases, ballgloves and also use helmets while batting.

In ice hockey, players use hockey sticks, pucks, two goals, ice skates, and helmets.

These two sports wear a certain amount of protection according to how rough the game is.

Ice hockey consists of a lot of physical contact and it is a very rough game.

Baseball, however, has less physical contact and is much safer to play.

Protective clothing are a must for ice hockey players, although baseball players do wear very little protection.

Even though ice hockey and baseball are almost two totlly different sports, they are both exciting and fun to play as well as watch.

A.4 Sample with rating 5

The Causes of My Coming to Gallaudet

This paper will discuss the following three causes of my coming to

Gallaudet: persuaded by high scholl teachers, passing Gallaudet's

Standard Academic Test, and major in psychology.

First of all, I was senior in high school and once thought about attending National Technical Institute for the Deaf in Rochester,

New York.

Several deaf teachers persuaded me to go Gallaudet University.

Because Gallaudet education is good for deaf and good social life.

The teachers knew that I would have great capability to get a job after graduating Gallaudet.

They were former Gallaudet.

They told me that Gallaudet would meet my need.

For example, there are deaf teachers and hearing teachers who sign, having experience in intership by Employmental Program Off-Campus, and deaf communities.

My teachers persuaded me to be on of Gallaudet alumni.

Secondly, I was curious to take Standard Academic Test of Gallaudet.

If I pass it as Preparatory, I would not go to the Preparatory school, so I would go to National Technical Institute for the Deaf.

If I pass it as freshman, I would go Gallaudet.

My teacher warned if receiving a thin letter that would say to thank for ???? and please try again, that means I was ?????.

If receiving a thick letter that had a lot of information about Gallaudet and application for attending school in fall.

It would say congradulations that I was passed.

When I received the thick letter in March, I was passed as a Freshman.

I was thrilled to be freshman in Gallaudet and decided to go Gallaudet certainly.

Last, I searched the majors what Gallaudet University has provided.

Gallaudet University is the Liberal Arts.

These majors are pyschology, physical education, computer programming, teachers, etc.

I was interested in psychology.

Psychology itself is general major and I could go to graduated in

Gallaudet University for specific field in psychology.

In conclusion, I decided to attend Gallaudet because I got a lot of persuasion from my high school teachers.

I would have good capability to be successful in the future.

Gallaudet University is great for deaf students.

Fortunately, I passed the test of Standard Academic Test if Gallaudet as Freshman.

I was interested in psychology major which Gallaudet University has provided.

How I am glad I am Gallaudet Student.

Appendix B

THE ERROR CODES

B.1 The full error set

Generic codes

`none`	No errors
`sp`	Spelling error
`un`	Unsure how to parse
`ran`	Random error
`nas`	Not a sentence

Determiners

`id`	Incorrect determiner	*I gave him* ***a cookies****.*
`md`	Missing determiner	*They may take* **___** *test.*
`ed`	Extra determiner	***The golf*** *is fun.*

Prepositions

`ip`	Incorrect preposition	***On*** *college there are nice people.*
`mp`	Missing preposition	*I went back* **___** *the dorm at ten.*
`ep`	Extra preposition	*There are many* ***of*** *men in the room.*

Modifiers

`ppp`	Prepositional phrase placement	***In the room*** *did he go?*
`ap`	Adjective placement	*I saw the* ***room blue****.*
`adv`	Adverb placement	*She is* ***being not*** *taught algebra.*
`aa`	Adjective/adverb confusion	*He speaks* ***clear****.* *The library is* ***quietly****.*
`adjf`	Adjective formation	*The class was* ***interest****.*
`ii`	Incorrect intensifier	***much*** *people*
`mi`	Missing intensifier	*I've never had so* **____** *fun.*
`ei`	Extra intensifier	*There are* ***much*** *many people here.*
`ht`	Here/there as a pronoun	***Here*** *has a lot to do.*

Noun phrases

`pl`	Pluralization problem (noun)	***Dog*** *followed me home.*
`pro`	Pronominalization error	*I like Sarah because* ***Sarah*** *is nice.*
`pc`	Pronoun case	*I like* ***she****.*
`pn`	Pronoun number	*I picked up the toys and put* ***it*** *away.*
`ipos`	Incorrect possessive	***his's*** *dog*
`mpos`	Missing possessive marking	*I like my* ***neighbor*** *cat.*
`nf`	Noun formation	*He is avoiding* ***responsible****.*

Agreement

`sv`	Subject/verb agreement	*The **dogs goes** outside.*
`so`	Subject/object agreement	*He is baseball **players**.*

Verb tenses

`ptf`	Past tense formation	*He **was teached**.*
`aux`	Auxiliary complement error	*She **is teach** the class.*
`inf`	Infinitive error	*I want ___ **go** home.*
`ger`	Gerund abuse	*I **working**.*
`act`	Activity	***Watch** golf is boring.*
`tc`	Tense context error	*I **become** more motivated.*

Verb phrase formation

`ia`	Incorrect auxiliary	*I **have** doing this.*
`ma`	Missing auxiliary	*I ___ **enjoying** the school.*
`ea`	Extra auxiliary	*I **do was** going home.*
`mv`	Missing verb	*He ___ a teacher.*
`ev`	Extra verb	*It **is have** a lot of buildings.*
`vpo`	Verb phrase order	*I **done have** this.*
`vf`	Verb formation	*He **thiefed** the book from the library.*
`pass`	Passive formation	*Students ___ required to study.*
`voi`	Voice	*The book **wrote**.*

Dropped focus elements

`ms`	Missing subject	*___ Got here yesterday.*
`mo`	Missing object	*I found ___.*
`ml`	Missing location	*I like the store because I shop ____ often.*

Dummy subjects and objects

`ids`	Incorrect dummy subject	***It** is a book on the table.*
`mds`	Missing dummy subject	*___ Lots to do at Gallaudet.*
`vo`	Verb/object agreement	*There **is people** in the room.*
`bh`	Be/have confusion	*There will **have** a meeting.*

B.1.11 Subordinate clauses

`irp`	Incorrect relative pronoun	*I saw the man* ***what*** *had a dog.*
`mrp`	Missing relative pronoun	*I saw the man* **____** *had a dog.*
`mw`	Missing "wh"	*I like it* **____** *people say hello.*
`erp`	Extra relative pronoun	***That*** *the man was going to the store.*
`mg`	Missing gap	*I saw the boy that she kissed* ***him****.*
`scf`	Subordinate clause formation	*I know what* ***do you*** *like.*
`oa`	Over-generalized agreement	*I saw John* ***kisses*** *Mary.*

B.1.12 Conjunctions

`ic`	Incorrect conjunction	*You can call me* ***that*** *I will be home.*
`mc`	Missing conjunction	*We went to the zoo* **___** *we saw a bear.*
`ec`	Extra conjunction	*I threw out my hat* ***and because*** *I sat on it.*

B.1.13 Miscellaneous

`sem`	Semantic word choice error	*I am* ***boring****.*
`syn`	Word choice error w/syntax problem	*I am* ***interesting*** *in botany.*
`neg`	Negation formation	*He* ***not*** *like dogs.*
`qu`	Question formation	*Why* ***he doesn't*** *like dogs?*
`cm`	Comparative problem	*He is* ***the best than*** *Ann at running.*
`idi`	Idiom problem	*I am not use cafeteria food.*
`asl`	ASL-ism	

B.2 The reduced error set

The following categories are the reduced sets of errors, grouped by syntactic structure, used in the factor analysis.

Sentence Structure: S V / S V O

```
nas   mv
ev    ms
mo
```

Sentence Structure: S V PP

```
ppp
```

Sentence Structure: S V adv

```
adv
ht
ml
```

Sentence Structure: delimiting

```
ic
mc
ec
```

Sentence Structure: questions

```
qu
```

Noun Phrases: Det N

```
id
md
ed
```

Noun Phrases: Prep NP

```
ip
mp
ep
```

Noun Phrases: Adj N

```
ap
```

Morphology: N (nom/acc/plural)

```
pl   pc
nf   so
```

Morphology: N +'s (possessive)

```
ipos
mpos
```

Morphology: modifiers

```
aa   adjf
ii   mi
ei
```

Morphology: V

```
vf
```

Morphology: V +s

```
sv
```

Morphology: V +ing

```
ger
act
```

Morphology: V +ed

```
ptf
```

Verb Phrases: Negations

```
neg
```

Verb Phrases: aux + V

```
aux   ia
ma    ea
vpo   pass
```

Verb Phrases: inf V

```
inf
```

Dummy Subjects

```
ids   mds
vo    bh
```

Subordinate Clauses

```
irp   mrp   mw
erp   mg    scf
oa
```

Comparisons

```
cm
```

Not used:

```
none   sp    un    ran
pro    pn    tc    voi
sem    syn   idi   asl
```

Appendix C

CORRELATIONS WITHIN HIGH-PROFICIENCY WRITING SAMPLES

The following chart shows the correlation matrix for the number of occurrences of the 76 error codes within the highest-quality set of writing samples (n=26). A bold ‘X’ in the table indicates that there was a correlation between the occurrences of the two error codes with $p<0.01$. A normal-weight ‘X’ indicates that $p<0.05$. All r values fall between 0.40 and 0.90.

BIBLIOGRAPHY

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

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

Berent, G., An assessment of syntactic abilities. In M. Strong (Ed.), *Language Learning and Deafness*. Cambridge: Cambridge University Press, 1988.

Berent, G., Learnability constraints on deaf learners' acquisition of English wh-questions. *Journal of Speech and Hearing Research*, 39:625-642, 1996.

Bochner, J., Error, anomaly, and variation in the English of deaf individuals. *Language and Speech*, 21(2):174-189, 1978.

Dulay, H. and Burt, M., Natural sequences in child second language acquisition. *Language Learning*, 24:37-53, 1974.

Gass, S., Language transfer and universal grammatical relations. *Language Learning*, 25(2):327-344, 1979.

Gass, S., A review of interlanguage syntax: language transfer and language universals. *Language Learning*, 34(2):115-132, 1984.

Gass, S. and Ard, J., L2 data: their relevance for language universals. *TESOL Quarterly*, 19:443-452, 1980.

Gould, S. J., *The Mismeasure of Man*. New York: W. W. Norton & Co., 1981.

Hawkins, J., Language universals in relation to acquisition and change: a tribute to Roman Jakobson. In *Current Issues in Linguistic Theory, v.49: new vistas in grammar: invariance and variation*.

Keenan, E. and Comrie, B., Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8(1):63-99, 1977.

Keenan, E. and Hawkins, S., The psychological validity of the accessibility hierarchy. In E. Keenan, *Universal Grammar: 15 Essays*. London: Croon Helm, 1987.

Lakshmanan, U., Functional categories and related mechanisms in child second language acquisition. In Flynn, S., Martohardjono, G, and O'Neil, W. (Eds.), *The Generative Study of Second Language Acquistion*. Mahwah, N.J.: Lawrence Erlbaum Assoc., 1998.

Lillo-Martin, D.C., The acquisition of English by deaf signers: is Universal Grammar involved? In Flynn, S., Martohardjono, G, and O'Neil, W. (Eds.), *The Generative Study of Second Language Acquistion*. Mahwah, N.J.: Lawrence Erlbaum Assoc., 1998.

McCoy, K. and Pennington, C., Generation issues in a second language learning tool. Generation issues in a second language learning tool. *Computational Linguistics*, 16:?, 1997.

McCoy, K., Pennington, C., & Suri, L., English Error Correction: A Syntactic User Model Based on Principled "Mal-Rule" Scoring. In *Proceedings of the Fifth International Conference on User Modeling* (UM1996), Kailua-Kona, Hawaii, 1996.

Michaud, L., Toward a Morphosyntactic User Model for Language Analysis and Generation. A PhD Proposal for the Department of Computer and Information Sciences, University of Delaware, 1999.

Michaud, L. & McCoy, K, Error Profiling: Toward a Model of English Acquisition for Deaf Learners. To appear in Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, Tolouse, France, 2001.

Michaud, L.; Stark, L.; and McCoy, K. Modeling the Acquisition of English: an Intelligent CALL Approach. To appear in *Proceedings of the Eighth International Conference on User Modeling* (UM2001), Sonthofen, Bavaria, 13-17 July 2001.

Quigley, S.P. and King, C., The language development of deaf children and youth. In S. Rosenberg (Ed.), *Handbook of Applied Linguistics*, 1992.

Quigley, S.P., Montanelli, D.S., and Wilbur, R.B., Some aspects of the verb system in the language of deaf students. *Journal of Speech and Hearing Research*, 19:536-550, 1976.

Quigley, S.P., Wilbur, R.B., and Montanelli, D.S., Question formation in the language of deaf students. *Journal of Speech and Hearing Research*, 17:699-713, 1974.

Quigley, S.P., Wilbur, R.B., and Montanelli, D.S., Complement structures in the language of deaf students. *Journal of Speech and Hearing Research*, 19:448-457, 1974.

Rutherford, W., Description and explanation in interlanguage syntax: state of the art. *Language Learning*, 34:127-155, 1984.

Schuster, E., The role of native grammars in correcting errors in second language learning. *Computational Intelligence*, 2:93-98, 1986.

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

Sleeman, D., Inferring (mal) rules from pupil's protocols. *Proceedings of ECAI-82*, 9:160-164, 1982.

Strong, M. *Language Learning and Deafness*. New York: Cambridge University Press, 1988.

Suri, L. and McCoy, K., A methodology for developing an error taxonomy for a computer assisted language earning tool for second language learners. Technical Report TR-93-16, Dept. of CIS, University of Delaware, 1993.

Swisher, V., The language-learning situation of deaf students. *TESOL Quarterly*, 23:239-257, 1989.

Tarallo, F. and Myhill, J., Interference and natural language processing in second language acquisiton. *Language Learning*, 33(1):55-76, 1983.

Tarone, E., Interlanguage as chameleon. *Language Learning*, 29:181-191, 1979.

Vainikka, A. and Young-Scholten, M., The initial state in the L2 acquisition of phrase structure. In Flynn, S., Martohardjono, G, and O'Neil, W. (Eds.), *The Generative Study of Second Language Acquistion*. Mahwah, N.J.: Lawrence Erlbaum Assoc., 1998.

Vygotsky, Lev. *Thought and Language*. The MIT Press, Cambridge, Massachusetts, 1986. Translation revised and edited by Alex Kozulin; originally published in 1934.

Weischedel, R. and Sondheimer, N., Meta-rules as a basis for processing ill-formed input. *American Journal of Computational Linguistics*, 9(3-4):161-176, 1983.

Weischedel, R., Voge, W., and James, M., An artificial intelligence approach to language instruction. *Artificial Intelligence*, 10:225-240, 1978.

Wilbur, R.B., An explanation of deaf children's difficulty with certain syntactic structures of English. *The Volta Review*, 79(80):85-92, 1977.

Wilbur, R.B., Montanelli, D.S., and Quigley, S.P., Pronominalization in the language of deaf students. *Journal of Speech and Hearing Research*, 19:120-140, 1976.