Correcting Discourse-Level Errors in a CALL System for Second Language Learners[1]

Linda Z. Suri & Kathleen F. McCoy

Technical Report No. 94-02
May, 1993

Linda Z. Suri & Kathleen F. McCoy
Department of Computer and Information Sciences
103 Smith Hall
University of Delaware
Newark, DE 19716 U.S.A.
suri@udel.edu, mccoy@udel.edu

[1] This research was supported in part by NSF Grant #IRI-9010112. Support was also provided by the Nemours Foundation.

Correcting Discourse-Level Errors in a CALL System for Second Language Learners

Linda Z. Suri and Kathleen F. McCoy

1 Introduction

In this paper, we examine the need for Computer Assisted Language Learning (CALL) systems to identify and correct errors at the discourse level. We are interested in the kind of CALL system that takes a user's written text as input, identifies errors in the text, and offers the student corrective information. Obviously, such a system must correct *sentence-level errors,* such as syntactic errors and inappropriate lexical choice. While difficult themselves, sentence-level errors can be detected and corrected by a system that parses text one sentence at a time, and searches for errors, using a standard English grammar augmented with appropriate syntactic error rules in conjunction with a set of rules for finding semantic anomalies.

However, there is another task that must be addressed by an effective CALL tool, and that is the correction of *discourse-level* errors. By discourse-level errors, we refer to errors which 1) influence the understanding/processing of subsequent text and/or 2) require the previous discourse context to be taken into account in order to be identified and/or corrected. The errors which fall into this class range in severity from making the text seem "choppy" or awkward to making the text incomprehensible (due to the reader's inability to resolve pronouns, for example). The identification and correction of discourse-level errors require additional discourse machinery.

A CALL system could be developed for use by a native speaker of the target language of the CALL system, or for use by a second language learner of the target language. The correction of discourse-level errors is important for both of these kinds of systems. The CALL system we are designing provides corrective information to native signers of American Sign Language (ASL) learning English as a second language. We have examined several classes of discourse-level errors which are common in the written English of ASL natives, and we propose that a *local focus* tracking mechanism will enable a CALL system to correct them [Sur92a, SM93a]. The term local focus refers to the item/object/concept which a sentence is most centrally about, within the discourse context in which the sentence occurs. In a coherent discourse, the focus does not randomly change from one sentence to the next, but generally changes in some well-specified ways, as we will discuss later.

Most previous local focus tracking work has been geared towards how to generate and process coherent standard written English text. While we are interested in this problem as well, we have been working on focus tracking from the perspective of correcting written English. We have been looking at how to use a focus tracking algorithm to correct discourse-level errors in the written English of ASL natives.

In fact, studying the role of focus in the correction of the written language of a second language learner may prove to be particularly interesting since it may lend insight into how languages function differently at the discourse level. We have concentrated on determining how focus can be used to flag inappropriate choice of forms of subsequent reference (e.g., use of a pronoun where a full noun phrase (NP) is required, use of an empty category (EC)[2] where a full NP or pronoun is required, or inappropriate pronoun choice). One particularly interesting class of errors involves the use of a zero NP in contexts where a zero NP would be appropriate in ASL, but not in English. Our particular CALL project may lead to a better understanding of how focus and subsequent reference function in English and in ASL.

In this paper, we will give examples of errors involving subsequent reference found in our analysis of writing samples from ASL natives. We will discuss how these errors can be explained by differences between ASL and English. These differences motivate the use of a *local focus tracking* (or focusing) algorithm, which will also be described. Finally, we will show how we can use focus tracking to identify and correct the illegal NP omissions, and to flag inappropriate pronominalization. This work should be applicable to the correction of reference errors in other CALL systems for second language learners.

[2] An *empty category* or *zero NP* is a noun phrase that is not signed or spoken but which must be present in the syntactic structure underlying the signed or spoken sentence.

2 Handling Errors Involving Subsequent Reference

The devices available to a user of a language to refer to items already introduced into the discourse are quite varied, and the rules for their appropriate use can be quite complicated. Notice that *the form* of subsequent reference may serve several purposes in the discourse in addition to picking out a referent (i.e., an object, event, etc.). For example, the variety of subsequent reference tools available in English serve as cohesive devices (e.g., to help indicate a continuance or a change in topic) in addition to allowing the user to pick out the appropriate referent. In [SM93a], we argue that the different roles that subsequent reference devices might play in two languages is a potential source of error for second language learners.

Languages commonly have three forms of subsequent reference: full NP's, pronominal NP's, and zero pronouns. Of course, both ASL and English have the feature of full noun phrase reference. In our discussion, we will concentrate on the use of pronominal reference and zero NP's (and thus implicitly discuss the use of full noun phrases). We will first discuss the feature of pronominal reference in ASL; that is *how* pronominal reference is achieved and *when* pronominal reference can/should be used in ASL. We will contrast this with a discussion of pronominalization in English. The differences between the languages in terms of when pronominal reference can/should be used allows us to explain some pronominalization errors we have found in the written English of ASL natives.

Then, we will discuss a kind of of zero pronominal reference that is used in ASL but not in English. Our claim is that the writer's use of a rule about when zero NP's can be used in ASL may explain some of the illegal NP omissions in the written English of ASL natives. We give examples of such omissions.

2.1 Pronoun Use in ASL

In ASL, a *locus* or location may be associated with an object, concept, event, etc., and then pronominal reference to that referent can be achieved by indicating the locus. A locus may be indicated by pointing or gazing at the locus, or by using that locus as the starting or end point of an inflecting verb.

For a referent that is present, the locus is the location of the referent itself. For referents which are not present, each referent may be associated with a locus in the signing space in front of the signer's body. "This is accomplished by producing the sign for the referent at some arbitrary locus in space, or making the sign and then pointing to the locus with the index finger, or by eyegaze in the direction of the locus while making the sign." (p. 25-6, [LM91]) Abstract loci associations persist in discourse until a new framework is established, and the number of such loci, while theoretically unlimited, typically does not exceed 5.

2.2 Pronoun Use in English

While it may or may not be apparent to a native speaker of English, the rules for when and when not to pronominalize a noun phrase in English are very complicated and dependent on several factors. These factors include the semantics of the items in the discourse, the semantics of the verb that an NP is an argument of, the focus history of discourse entities and information about focus movement, the intended focus of entities in the current sentence, pragmatics, world knowledge, and knowledge shared by the discourse participants.

English has a small number of pronouns and these carry some semantic information about their referent (e.g., gender and number). Pronoun use in English is largely dependent on focus or what has been made prominent in previous discourse. The use of a pronoun is often an indicator of continued prominence of the referent. However, there are some situations where a pronoun cannot be effectively used even though focus rules might favor a pronoun. These are cases of potential ambiguity resulting from the presence of other items in reasonably high focus in the discourse with the same gender and number features.

2.3 Differences between Pronominal Reference in ASL and English

In some sense, the use of a locus (for pronominal reference) in ASL is analogous to using pronouns in English because both are forms of reduced reference. However, the similarity stops there. Perhaps the most important difference is that in ASL pronominal reference is the normal form of reference to an item that has already been mentioned. Lexical repetition is used in a very limited fashion in ASL and when it is used it seems to serve the function of reaffirming the present topic (p. 235, [WP83]). In English, several factors influence whether or not something is referred to pronominally. For instance, the use (and comprehension of) pronominalization is highly dependent on whether or not (and to what extent) that item has been in focus in the preceding discourse. In fact, in English, the use of a pronoun (as opposed to lexical repetition) frequently serves the function of reaffirming the present topic.

2.4 Explanations for Pronoun Errors

Our analysis of writing samples collected from deaf (American) writers uncovered pronominalization errors. Many of these errors can be attributed to the differences in the way that pronominalization is used in ASL and in English. Below we discuss two excerpts with pronominalization errors.[3]

1. "Fraternities and Sororities here at XYZ DO provide [a] social life. Some examples: parties; gettogethers; workplaces; IM; and sports. At parties that *they* host..."

   Possible correction: "At parties that *the fraternities and sororities* host..."

In this excerpt, the pronoun "they" in the final sentence could not refer to anything more recent than "Fraternities and Sororities", based on semantics (people or organizations host parties; parties, gettogethers, workplaces, IM or sports do not host parties). Nevertheless, the pronoun resolution of "they" is difficult for a native speaker of English because the focus of the discourse has shifted away from fraternities and sororities to various kinds of social events. Notice that in ASL this shift would not preclude the use of a "pronoun" (locus) to refer to fraternities and sororities in this instance. Thus, a possible explanation of this error is that the writer does not understand the relationship between focus and pronoun use in English since that relationship is not present (or is very different) in ASL.

2. "Abolishing fraternities and sororities at XYZ would be a very unwise thing to do. If so, no fun; no unity; and no social life would result. Being a Greek is what every student dreams of when they enter college. If we abolish *them*, that dream will never come true. Not only should Greek organizations stay open due to traditional reasons, *it* offers a wonderful bond between [...] brothers and sisters. For an example, my close friend, Amy, is a member of a sorority here at XYZ. I have seen the closeness she and her fellow sisters [have] and I envy that. Joining an organization gives you wonderful personal growth. *They* can help you improve in your areas of weaknesses."

In this excerpt, we have italicized three pronouns which have been used incorrectly according to the intuitions of a native speaker of English. We will consider each of them in turn. The first error involves the *them* in "If we abolish *them*..." The most likely correction for this is "If we abolish *fraternities and sororities*," which would be consistent with what the essay is about. This error is analogous to the error described in the previous excerpt. In particular, the writer doesn't recognize that *fraternities and sororities* are no longer prominent enough to be referred to pronominally at this point in the discourse.

The second error involves the *it* in "... *it* offers a wonderful bond..." This error can be explained in two different ways, depending on what the writer is trying to refer to. The first possibility is that the writer has made a pronoun choice ("it" for *Greek organizations*) which is incorrect since "it" and *Greek organizations* do not agree in number. The second is that the writer is trying to refer to *being a Greek* with the pronoun "it". This would be an incorrect usage since the notion of *being a Greek* isn't prominent enough to be referred to pronominally (in English) at this point in the discourse.

For the third error ("*they* can help you"), there are several things the writer could have intended to refer to, but each of these is problematic. For example, the most recent NP that could co-refer with "they" in terms of number agreement is *she and her fellow sisters*. However, the sentence would not make sense pragmatically if that were the intended referent. On the other hand, the writer could have intended to refer to *the fraternities and sororities*, *(Greek) organizations*, or *the members of fraternities and sororities*). However, these items are fairly far back in the discourse and no longer prominent enough for pronominal reference. Alternatively, the writer could have intended to refer to *joining an organization* or *being a Greek*. However, then there is a number feature mismatch between the intended referent and the pronoun.

In the above examples, there are three possible causes for errors involving pronoun use. The first is that the writer does not appreciate that in English the form (i.e., number and gender) of pronoun may not

[3] For clarity, we have italicized the words of a sentence which are related to errors being discussed. We have sometimes corrected portions of a sentence, placing the corrections in square-brackets, [ ], when the lack of correction of those portions would have made it difficult to understand an error being discussed. Also, to prevent the identification of writers, we have changed the names of institutions (to "XYZ college" or "XYZ", for example).

be enough to fully specify its referent, although in ASL a referent is uniquely identified by the locus of the pronominal reference. The second is the that writer does not appreciate the role of focus in pronominalization, and thus he or she has tried to refer pronominally to an NP that is no longer prominent enough to be referred to pronominally. The third is that the writer does not recognize a mismatch between the number and/or gender of the intended referent and the number and/or gender of the pronoun. This could result in an incorrect choice of a pronoun.

These sources of error can be explained by differences in the nature of pronominalization (since pronominal reference uniquely identifies a referent in ASL but not in English), and by differences between the rules of when to use pronominalization in ASL and English. Errors involving pronoun use are likely to persist until the writer learns the relationship between focus and pronominal use in English, in addition to understanding how/what semantic features are captured by English pronouns.

2.5 Empty Categories in ASL and English

Another form of subsequent reference used in ASL involves the use of Empty Categories or zero NP's. Recall that an empty category or a zero NP is a noun phrase that is not signed or spoken but which must be present in the syntactic structure underlying the signed or spoken sentence. Here we concentrate on the EC's that occur with *plain verbs* in ASL [LM91]. The plain verbs of ASL have no morphology that would help the "hearer" recover an implicit referent. However, these verbs do allow null arguments (and, thus, empty categories) in some contexts. Lillo-Martin [LM91] argues that the deletions with respect to these ASL verbs are similar to deletions in languages (like Chinese) which are termed *discourse-oriented* languages. In these languages, the deleted NP can be recovered from context. Her analysis of deletions/EC's with respect to plain verbs in ASL is based on Huang's analysis of EC's in Chinese. Specifically, Lillo-Martin argues that ASL allows *Topic NP Deletion*, i.e., for the topic of a sentence to be "deleted under identity with a topic of a preceding sentence" [Hua84].

Although English does have some forms of empty categories, it does not have the feature of allowing Topic NP Deletion, and so it does not have the corresponding zero NP's. In sum, ASL has zero NP's that arise from the deletion of an NP that co-refers with the topic of a previous sentence (or a discourse topic), but English does not.

2.6 An Explanation for NP Omission Errors

Consider the following examples of English text (written by deaf American writers) with NP omissions:

1. "I think that XYZ College should require all deaf students to take speech and speechreading courses. Therefore, they can improve their oral skills for their future use. I am going to tell you [...] why the deaf student should take _."

2. "There are many things I like about XYZ. They offer supporting services like interpreters and notetakers for mainstream classes which I had experiences [with in] my public schools. Now XYZ/PDQ offers [the] same thing that my school offered but [... with] better supporting services. That is [what] I like about XYZ.

   But [the] one thing [that] worries me [the] most about XYZ/PDQ is financial problems. I hope I could find some ways to solve _."

3. "First, in summer I live at home with my parents. I can budget money easily. I did not spend [a] lot of money at home because at home we have [a] lot of good foods, I ate [a] lot of foods. While living at college I spend [a] lot of money because _ go out to eat almost everyday. At home, sometimes my parents gave me some money right away when I need _. While in college, I could not ask my parents for money right away because I live in XYZ and my parents live in QRS. It is too far."

Each of the above examples is missing (one or more) noun phrases and, as a result, the texts are ungrammatical in English. Informally, we can see that each of the deletions in the above excerpts can be explained as resulting from the writer's application of the rule of Topic NP Deletion. For example, in the last sentence of the first excerpt, it seems the writer has deleted a referent to *speech and speech reading courses* since the courses are the topic of the first sentence (and perhaps the second sentence, if "they" is taken to refer to *speech and speech reading courses* and not to *deaf students*, although we prefer the latter reading). In the second example, it seems the writer has deleted a referent to *financial problems* since *financial problems* is the topic of the previous sentence.

For each of the above examples, discussions with ASL informants indicated that the corresponding ASL discourse would be acceptable/grammatical if the omitted NP were not signed, pronominally referenced, or indicated by verb agreement. Our informants reported that the omitted items would be understood from the context. This supports our proposal that these errors can be explained by the use of the Topic NP Deletion rule.

3 Local Focus, Zero NP's, and Pronominalization

In the case of illegally omitted NP's and several of the pronominal reference errors noted above, a CALL system would benefit from an algorithm that tracks the local focus of the discourse. As noted earlier, in a coherent discourse, the focus does not randomly change from one sentence to the next, but generally changes in some well-specified ways. Thus, focus tracking algorithms must record items that can potentially be the focus of subsequent discourse, and must determine when and how the focus shifted.

A primary use of local focus is in resolving anaphors, particularly pronouns and (in our work) zero NP's arising from illegally omitted noun phrases. Because pronouns carry relatively little information about their referents, one would expect a pronoun to co-specify a focused element. Our focusing algorithm uses data structures to store items that can be focused in subsequent discourse. Our pronoun resolution algorithms impose preferences among the data structures to suggest the likely referent of a pronoun. Like other focusing research (e.g., [Sid83, Car87], our algorithms rely on other components that use semantics, pragmatics and general knowledge inferencing to accept or reject these proposed referents. The purpose of a focusing algorithm is thus to limit the amount of inferencing that is required in order to find the antecedent of an anaphor[4]: it is much easier to confirm a co-specification using inferencing, than to suggest and find a co-specification using inferencing. Thus, a measure of a good focusing algorithm is how well it predicts the likely co-specifier of a pronoun or zero NP.

Once all anaphors in the current sentence are resolved, the focusing data structures must be updated to represent preferences about what will be discussed (and what is most likely to be pronominalized) in the subsequent discourse. Thus, pronoun resolution and focusing are mutually dependent processes.

3.1 Focusing Algorithm Overview

We will not give the details of our focusing algorithm here. Instead, we will give an overview. Our focusing algorithm is based on Sidner's local focusing algorithm [Sid83], and as such we track both a *current focus* (or local focus or topic), and an *actor focus*. Since we will be discussing only omissions of non-agent NP's, we can restrict our description to the tracking of the current focus.[5]

The current focus of a sentence is that item/object/concept which the sentence is most centrally about. This item is stored in the current focus or CF data structure. All other NP's in the sentence, and the VP, are stored in the potential focus list or PFL. Except for certain kinds of sentences (such as It-clefts or There-insertion sentences), it is difficult for a reader to determine the focus of a discourse-initial sentence, so our algorithm picks an *expected* focus for the first sentence based on syntax. At any given point in a well-formed text, after the first sentence, the writer has a number of options. (We list these options in the order of likelihood, and thus the order of preference used by our focusing algorithm.) The writer can:

- Continue talking about the same thing; in this case, the CF doesn't change.
- Talk about something just introduced; in this case, the CF is selected from the previous sentence's PFL.
- Return to a topic of previous discussion; in this case, that topic must have been the CF of a previous sentence.
- Discuss an item previously introduced, but which was not the topic of previous discussion; in this case, that item must have been on the PFL of a previous sentence.

[4] Focusing and inferencing is necessary for resolving full NP's as well as zeros and pronouns. However, we limit our discussion to pronoun and EC resolution.

[5] A comparison of the focusing framework and the centering framework (an alternative to the focusing framework introduced by Sidner) can be found in [SM93b], and further discussion of focusing (with comments on centering) can be found in [SM93c].

The decision (by the reader/hearer/algorithm) as to which of these alternatives was most likely chosen by the speaker (or which element of the current sentence is the new CF) is based on many factors. These factors include the syntactic roles of anaphora of the current sentence, the type of each anaphor (full NP, pronominal NP, or illegal zero) and the focus history of the discourse (in particular whether the co-specification of each anaphor is the CF, a previous CF, or a member of the current PFL or a previous PFL). It is these just-mentioned factors that our focusing algorithm tries to capture and utilize in computing the current focus and in computing the antecedents of anaphors. Again, confirmation of suggested anaphora co-specifications requires inferencing by other components based on general knowledge, pragmatics and semantics.

At each sentence in the discourse, the CF and PFL of the previous sentence are stacked for the possibility of subsequent return.[6] When one of these items is returned to, the stacked CF's and PFL's above it are popped, and are thus no longer available for return.

3.2 Filling in Illegally Omitted NP's

Just as the fact that pronouns carry relatively little information suggest that their co-specifiers must be highly focused, we would expect that the co-specifier of a zero NP should also be highly focused. Furthermore, because we believe that the illegally omitted NP's in the written English of ASL natives are the result of the writer deleting an NP under the rule of Topic NP deletion, we expect that the co-specifier of an illegally omitted NP should be a previous topic. Thus, to identify the co-specifier of an omitted NP, our algorithm imposes preferences among the contents of the focusing data structures so as to prefer highly focused elements.

More specifically, when we encounter an omitted NP that is not the (deep) subject of the sentence, we first try to fill the deleted NP with the CF of the immediately preceding sentence. If syntax or inferencing based on general knowledge, semantics, and pragmatics cause this co-specification to be rejected, we then consider members of the PFL of the previous sentence as fillers for the deleted NP. If these too are rejected, we consider stacked CF's and elements of stacked PFL's.

Below, we will describe the behavior of our algorithm on an example from our collected texts containing a deleted non-subject NP. One of the major contributions of our focusing work is addressing how focus should be computed for complex sentences [SM93c]; this problem has not been explicitly addressed in previous focusing work. Although we have not explained the details of processing complex sentences here, the demonstration of the algorithm reflects our extensions through the ordering of elements on the PFL.

> *"(S1) I think that XYZ should require all deaf students to take speech and speechreading courses.*
> *(S2) Therefore, they can improve their oral skills for their future use.*
> *(S3) I am going to tell you that why the deaf student should take _"*

After S1, the CF is `SPEECH AND SPEECHREADING COURSES (S & SR COURSES)`; the PFL contains the `TAKE S & SR COURSES` VP, `DEAF STUDENTS`, the `REQUIRE` VP, `XYZ`, the `THINK` VP, `I`; and the CF stack and PFL stack are empty.

After S2, the CF is `DEAF STUDENTS`[7]; the PFL contains `THEIR ORAL SKILLS`, `THEIR FUTURE USE`, and the `CAN IMPROVE` VP; the CF stack contains `S & SR COURSES`; and the PFL stack contains the contents of the last PFL.

When we process S3, in order to fill in the missing NP, our algorithm first tries to fill this (non-subject) NP with the CF. The CF is presently `DEAF STUDENTS`. This possibility is rejected under our *Consistent Deletion* constraint, which says that if one NP is deleted under coreference with a previous topic, then all NP's in that clause the co-specify that NP would also have been deleted under co-reference with a previous topic (and thus no deleted NP can co-specify another non-deleted NP in the same clause). Next, our algorithm will try to fill the missing NP with elements of the PFL. However, we will reject `THEIR ORAL SKILLS` and `THEIR FUTURE USE` by semantics, and the VP will be rejected because we can not fill an NP with a VP. Then, the algorithm will try to fill the missing NP with the first item on the CF Stack. This is `S & SR COURSES`, which is acceptable according the semantics, pragmatics, and general knowledge, and thus the NP is filled with this referent. This result matched our intuitions for what the writer intended.

[6] Sidner did not stack PFL's. We have identified samples of discourse that require stacking the PFL's.

[7] This analysis assumes that the "they" in S2 co-specifies `DEAF STUDENTS`; however, the analysis for the case that "they" co-specifies `S & SR COURSES` has the same end-result for filling in the omitted NP.

3.3 Correcting Pronominalization Errors

Focusing can also be used to flag errors involving inappropriate pronoun use. In particular, if our pronoun resolution algorithms can not find a co-specifier among highly focused entities, and there is no indication that the focus has been popped (e.g., the start of a new paragraph or the use of clue words), then the algorithm would indicate inappropriate pronominalization. In fact, whether or not the item stored in a particular focusing data structure is prominent enough to serve as the co-specifier of a pronoun is partly dependent on focusing expectations. (For example, if one item remains the focus for a long time, and then the focus shifts, it is likely that this focus shift involved a parenthetical comment. This expectation would not hold if the focus has been frequently shifting.)

Note that this approach will allow us to identify that there is a problem with the use of a pronoun, but it still leaves open whether the problem results from the writer using a pronoun to refer to an item which is no longer prominent enough to be pronominally referenced, or using a pronoun that does not match the intended referent in number and/or gender. However, the system could offer the CALL user possible referents (some being no longer prominent entities, others being entities that do not match the pronoun in gender and/or number), and ask the user what the intended referent was.[8] Once the intended referent is established, the system could give appropriate corrective information based on information stored by the focusing algorithm.

3.4 Relationship of our work to other Focusing Algorithms

The two most prominent frameworks for local focusing are the focusing framework established by Sidner ([Sid83]) and centering ([GJW83]). One important contribution of our work is explicitly addressing how to handle certain classes of complex sentences; the problem of how to process complex sentences has not been explicitly addressed by previous focusing research.[9] Because complex sentences are prevalent in standard written English, it is imperative that one determine how to appropriately process them. We have begun to address how to resolve pronouns and compute the focus in particular kinds of complex sentences, e.g., sentences involving sentential clauses conjoined by "because".[10]

Also, based on analyses of standard written English, we have found the need to modify Sidner's focusing algorithm [Sur92a] and to modify her pronoun resolution algorithms. In addition, Sidner did not address the resolution of zero NP's (either legal zero NP's in English, or illegal ones of the nature that we have found among ASL signers acquiring English).

Grosz, Joshi and Weinstein [GJW83] proposed a local focusing algorithm that they refer to as a *centering* algorithm, and used by [BFP87, Kam86, Gun80, Wal89]). Centering does not record the local focus history, and thus it assumes that the co-specifier of a pronoun can be found in the preceding or current sentence. However, we have identified instances of naturally occurring text where the co-specifier of a pronoun is not in the immediately preceding sentence, but a few sentences back (and recorded in our local focusing algorithm data structures); and yet we found no indication in these texts that the co-specifier should be found via global focusing. While the notion of centering has been assumed to be different than that of local focusing (e.g., focusing keeps an actor focus and a local focus while centering keeps a center and a preferred center), our work has shown that the frameworks are more similar than often assumed. It is the case, however, that the algorithms make different predictions in certain circumstances. Part of our ongoing research involves clarifying these differences and arguing why focusing is a more appropriate framework. (See [SM93b] for a discussion of the two frameworks, and [SM93c] for a discussion of our work on handling complex sentences.)

4 Conclusion

If a person is to gain proficiency in a language, not only must he or she have appropriate command of individual syntactic structures and lexical items, but he or she must learn to use them appropriately *in*

[8] A human tutor would often face the same predicament in deciding which item was the intended referent, though rich pragmatic knowledge may sometimes enable a human tutor to infer the referent without consulting the writer.

[9] Carter [Car87] briefly discussed the interaction of intrasentential and intersentential anaphora, but he did not address complex sentences. [Wal89] indicated that centering needs to handle multiple subjects, but did not specify how to do that.

[10] For some of these sentence types the focusing preferences may be dependent on genre; for example, whether elements of the *cause* or the *effect* clause of a statement conjoined by "because" are most often the focus of subsequent discourse may be dependent on whether the writing is about a science topic or a trip to the beach.

context. Failure to use syntactic structures and lexical items appropriately results in discourse-level errors. Such errors influence the understanding/processing of subsequent text and/or require the previous discourse context to be taken into account in order to be identified and/or corrected. Our analysis of second language writing uncovered a large number of discourse-level errors [Sur92b, SM93a]. Thus, the correction by CALL systems of discourse-level errors is important. The identification and correction of discourse-level errors requires the use of discourse information and discourse mechanisms.

This paper has identified one kind of information that is useful for correcting several kinds of discourse-level errors: local focus. We have discussed how to use a focusing algorithm to correct illegal NP omissions in the written English of ASL natives. This novel use of a focusing algorithm is well-founded in terms of ASL linguistics [BC80, KB79] and null-argument structures in ASL [LM91]. Since other languages (e.g., Chinese) share Topic-NP deletion in common with ASL, this work should be applicable to the correction of English written by speakers of other languages. We have also discussed how a local focusing algorithm could be used to handle errors involving inappropriate pronominalization. Again, our work on these errors should be helpful to second language learners from a variety of first language backgrounds.

Furthermore, we have identified that focusing algorithms need to address how pronouns should be resolved and how focus should be computed in complex sentences [Sur92a], and we have started to address this problem [SM93c].

In our current research, we are continuing to investigate how focus and other discourse information can be used to improve the effectiveness of CALL systems.

References

[BC80] C. Baker and D. Cokely. *American Sign Language: A Teacher's Resource Text on Grammar and Culture.* TJ Publishers, Silver Spring, MD, 1980.

[BFP87] Susan E. Brennan, Marilyn W. Friedman, and Carl J. Pollard. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, pages 155–162, 1987.

[Car87] David Carter. *Interpreting Anaphors in Natural Language Texts.* John Wiley and Sons, New York, NY, 1987.

[GJW83] Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, pages 44–50, Cambridge, Mass, June 1983.

[Gun80] Jeanette K. Gundel. Zero NP-anaphora in Russian: A case of topic prominence. In *Papers from the Parasession on Pronouns and Anaphora*, pages 139–146. Chicago Linguistic Society, University of Chicago, 1980.

[Hua84] C.-T. James Huang. On the distribution and reference of empty pronouns. *Linguistic Inquiry*, 15(4):531–574, Fall 1984.

[Kam86] M. Kameyama. A property-sharing constraint in centering. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, pages 200–206, Cambridge, Mass, 1986.

[KB79] E. S. Klima and U. Bellugi. *The Signs of Language.* Harvard University Press, Cambridge, MA, 1979.

[LM91] Diane C. Lillo-Martin. *Universal Grammar and American Sign Language.* Kluwer Academic Publishers, Boston, 1991.

[Sid83] Candace L. Sidner. Focusing in the comprehension of definite anaphora. In Robert C. Berwick and Michael Brady, editors, *Computational Models of Discourse*, chapter 5, pages 267–330. MIT Press, Cambridge, MA, 1983.

[SM93a] Linda Z. Suri and Kathleen F. McCoy. A methodology for developing an error taxonomy for a computer assisted language learning tool for second language learners. Technical Report TR-93-16, Dept. of CIS, University of Delaware, 1993. (Submitted for journal publication; draft available as technical report).

[SM93b] Linda Z. Suri and Kathleen F. McCoy. *Comparing Local Focusing and Centering.* In the proceedings of the Workshop on Centering Theory in Naturally-Occurring Discourse, Institute for Research in Cognitive Science, University of Pennsylvania, May, 1993.

[SM93c] Linda Z. Suri and Kathleen F. McCoy. *Focusing and Pronoun Resolution in Particular Kinds of Complex Sentences.* In the proceedings of the Workshop on Centering Theory in Naturally-Occurring Discourse, Institute for Research in Cognitive Science, University of Pennsylvania, May, 1993.

[Sur92a] Linda Z. Suri. Correcting illegal NP omissions using local focus. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics (Student Session)*, pages 273–275, June 1992.

[Sur92b] Linda Z. Suri. Using local focus to correct illegal NP omissions (a Ph.D. proposal). Technical Report TR-93-07, Dept. of CIS, University of Delaware, 1992.

[Wal89] Marilyn A. Walker. Evaluating discourse processing algorithms. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, pages 251–261, 1989.

[WP83] R. Wilbur and L. Petitto. Discourse structure in American Sign Language (or, how to know a conversation when you see one). *Discourse Processes*, 6:225–241, 1983.