A Methodology for Extending Focusing Frameworks

Linda Z. Suri and Kathleen F. McCoy
University of Delaware
Newark, DE 19716
LZS@CS.WUSTL.EDU, MCCOY@CIS.UDEL.EDU

Abstract

We address the problem of how to develop and assess algorithms for tracking local focus and for proposing referents of pronouns. We note that other focusing research has not adequately addressed the processing of complex sentences. We discuss issues involved in processing complex sentences and we review a methodology used by other researchers to develop their focusing frameworks. We identify difficulties with that methodology and difficulties with using a corpus analysis to extend focusing frameworks to handle complex sentences. We describe our methodology for extending focusing frameworks to handle complex sentences and we explain how it overcomes the difficulties faced by the other approaches.

Introduction

The central problem addressed in this work[1] is how to develop and assess algorithms for tracking local focus and for proposing referents of pronouns for use in Natural Language Processing (NLP) systems. By *local focus*, we refer to the person, object, property or concept that a sentence is most centrally about within the discourse context in which it occurs. The appropriate movement and marking of local focus, and the appropriate choice of the form of a Noun Phrase (NP) based on local focus information, are considered to contribute to the local coherence exhibited by discourse ((Sidner 1979), (Grosz, Joshi, & Weinstein 1983), (Carter 1987), and others).

In addition, local focus information is one source of information that is used by readers and hearers for interpreting pronouns. In fact, local focus tracking and pronoun resolution are mutually dependent processes. The local focus information influences pronoun resolution, and pronoun resolution, in turn, influences updating focus information. Therefore, the tracking of local focus is crucial for the interpretation of pronouns.

[1]This research was supported by NSF Grant #IRI-9010112, the Nemours Foundation and a Unidel Summer Research Fellowship from the Department of Computer and Information Sciences at the University of Delaware.

In this work we explore a robust mechanism for assessing and extending local focusing algorithms. After identifying flaws with some alternative approaches, we introduce a two part methodology which we term the Semantically-Slanted Discourse (SSD) Methodology. The first part of the methodology consists of an exploratory phase in which possible extensions to a focusing algorithm are discovered through the use of carefully constructed discourses which rely on the potential tension between focusing and world knowledge factors in pronoun resolution. We explain that this phase will then be followed by a corpus analysis to confirm its findings.

In this paper we briefly introduce the notion of local focusing and what a local focusing algorithm is intended to capture. Other work on focusing has not adequately addressed the processing of complex (i.e., multi-clausal) sentences. We discuss a number of issues involved in their processing in order to motivate the need for a methodology such as we describe. We review a methodology used by other researchers to develop and extend their focusing frameworks, and we identify some difficulties with that methodology. We examine the possibility of using a corpus analysis to extend a focusing framework, and briefly describe potential problems with such an approach. Finally, we describe our Semantically-Slanted Discourse Methodology, and show how it can be used to extend a local focusing framework to handle a type of complex sentence.

What is local focusing?

We use the term *local focusing framework* to refer to a theory or framework consisting of a (set of) focus tracking algorithm(s) and a (set of) pronoun resolution algorithm(s). A local focusing framework records and makes use of information about *focusing factors* and indicates how these factors influence pronoun resolution. Generally, focusing factors include such things as:

1. grammatical role – in several focusing frameworks, some grammatical roles (e.g., surface subject and surface direct object) are considered indicative of

what is in focus. Also, a pronoun resolution algorithm might prefer to find an antecedent which has the same grammatical role as the pronoun.

2. pronoun use - most, if not all, focusing frameworks are constructed on the basis of the belief that pronominalization is often indicative of focus.
3. constancy of focus - many focusing frameworks assume that if an item that was the focus of the previous sentence occurs in the current sentence, it is more likely to be focused in the current sentence.
4. clue words and phrases - most, if not all, focusing research takes into consideration that clue words and phrases might affect pronoun resolution and what is focused in a sentence.
5. syntactic form - in some frameworks, certain syntactic forms (e.g., clefted sentences) might indicate an item as highly focused.

The specific list of focusing factors and the way they interact are different for different focusing frameworks. A particular focusing framework must identify a set of focusing factors and must indicate how these focusing factors interact to suggest co-specifications for pronoun resolution and to identify the focus of a sentence.

A local focusing framework is not intended to *independently* interpret pronouns. Rather, a local focusing framework is intended to suggest co-specifications for pronouns in a reasonable order. An inferencing mechanism that makes use of *semantic factors* (such as semantic case constraints, world knowledge, rhetorical relations, etc.) must be used to confirm or reject a suggested co-specification. Thus, local focusing frameworks are intended to capture a coherence factor in discourse which influences preferences for how to resolve pronouns independent of semantic factors.

Processing Complex Sentences: A Reason for Extending Focusing Algorithms

Although complex sentences are prevalent in written English, other local focusing research (Focusing: (Sidner 1979), (Carter 1987); centering: (Grosz, Joshi, & Weinstein 1983), (Brennan, Friedman, & Pollard 1987), (Walker 1989), (Kameyama 1986), (Walker, Iida, & Cote 1992), (Brennan 1993), (Kameyama 1993), (Kameyama, Passonneau, & Poesio 1993), (Linson 1993), (Hoffman & Turan 1993), (Walker 1993); and PUNDIT: (Dahl 1986), (Palmer *et al.* 1986), (Dahl & Ball 1990)) did not explicitly and/or adequately address how to process complex sentences. Thus, there is a need to extend focusing algorithms.

Notice that there are a number of ways that a given type of complex sentence might be handled. For instance, consider processing a complex sentence of the form "SX because SY," where SX and SY each consist of a single clause. One might imagine processing the SX clause and then the SY clause (i.e., resolving the pronouns in these clauses and updating the focusing data structures) as if the clauses were a sequence of simple (i.e., single clause) sentences. On the other hand, it may be the case that for this type of complex sentence, the sentence should be treated as a single unit of processing with elements of one of the clauses dominating the processing. (For further discussion of these and other possible processing possibilities, see (Suri 1993).)

The question we address is how one can appropriately extend a focusing mechanism to handle various kinds of complex sentences.[2]

A Methodology Used in Other Local Focusing Work

Recall that local focusing theories are attempting to capture patterns of focus movement and patterns of relations between anaphors and their antecedents that are independent of semantics, world knowledge, rhetorical relations, etc. Because of this, the method for determining how to process particular kinds of complex sentences that might seem the most natural is to construct semantically-neutral discourses[3] that involve the type of complex sentence under study, and gather linguistic judgments to determine how people prefer to resolve the pronouns. In fact, in exploring *other aspects* of local focusing frameworks, other literature appears to have tried to make use of semantically-neutral texts in this fashion (e.g., (Brennan, Friedman, & Pollard 1987), (Walker, Iida, & Cote 1992)).

However, in trying to construct discourses to determine how to process a particular kind of complex sentence, we realized it is difficult to construct discourses that are truly semantically-neutral *and* sound natural. This task is further complicated by the need to construct a number of semantically-neutral texts in order to control for and isolate each of the factors that might affect how readers prefer to resolve pronouns.[4] These factors include the influence of the other complex sentence structures in the discourse, and the factors that affect focus computation and pronoun resolution for simple sentences, i.e., focus history, the syntactic roles of pronouns and their potential antecedents, verb aspect and tense, etc.

More importantly, when one is constructing texts *without the benefit of a systematic methodology*, one cannot be sure that the collection of constructed texts are representative of naturally-occurring text in terms of the interactive relationships within and across focusing and semantic factors, and their influence on pro-

[2]Note, it is possible that various types of complex sentences would each need to be handled differently.

[3]I.e., discourses whose pronouns cannot be unambiguously resolved on the basis of semantic/world knowledge factors alone.

[4]This need was not addressed by previous focusing work in an adequate or systematic fashion.

noun interpretation. As a result, there is a danger of tuning a theory to handle a discourse phenomena that is the exception rather then the norm in naturally-occurring situations.

Using a Corpus Analysis to Extend Frameworks

Because of the problems associated with using constructed discourses, it is natural to turn to some kind of corpus analysis to extend a focusing framework. For example, one might measure how well an extension of a framework handles a type of complex sentence by measuring how accurately and efficiently it suggests referents for pronouns in texts which contain the type of complex sentence under consideration. One could count how often the extended framework suggests a wrong referent (which would not be rejected by an ideal inferencing mechanism), and how many referents it suggests (on average) to the inferencing component before the correct referent is selected.

In using this type of approach, one is faced with several potential difficulties. First, a focusing framework is intended to capture a reader's preferences for focus movement and pronoun resolution *independent* of world knowledge, semantics, and other pragmatic factors. Very large amounts of text would have to be analyzed to control the influence of these factors, yet, since there are few tools available for this type of analysis, this task would be formidable.

Second, a corpus analysis may be useful in *comparing* various extensions, but it is up to the designer to decide *which* extensions to compare. Using this approach, a novel extension cannot emerge by becoming evident as a side effect of the analysis. Generally speaking, before a corpus analysis can be used, the researcher must have made all decisions concerning the processing. That is, all possible extensions of the framework (which are to be tested in the corpus analysis) must be completely identified. Notice, however, that there is no guarantee that the *correct* extension will be specified and tested. One might overlook the appropriate answers to how to segment sentences and how to process a particular kind of complex sentence. In addition, the number of possible extensions is likely quite large, and thus the number of alternative corpus analyses will likely be quite prohibitive.

Perhaps the most significant and problematic obstacle for determining how to extend a focusing framework to handle a particular kind of complex sentence via a corpus analysis is the following: if one does not know how to process many types of complex sentences, it is difficult to perform a corpus analysis to determine how to process a given type of complex sentence, since instances of that type of complex sentence are likely to be preceded and followed by other types of complex sentences. Furthermore, many sentences in the corpus are likely to involve multiple levels of complexity. Thus there is no way to isolate the influence of one complex sentence structure from the effect of another complex sentence structure.

Our Two-Part Methodology for Determining How to Process Complex Sentences

As we have pointed out, there are several potential problems with analyses using constructed discourses and with corpus analyses. Our methodology combines specific instances of both of these methodologies; these specific instances were designed to overcome the difficulties we identified.

The first part of our methodology involves systematically constructing discourses (of a type to be described) and gathering acceptability judgments on these discourses. The discourses are constructed in such a way so as to help identify a plausible extension of a focusing algorithm which would handle the type of complex sentence in question. The resulting extension must then be confirmed by a corpus analysis to ensure that the constructed discourses uncovered influences actually found in naturally-occurring text.

In the remainder of this paper, we focus on the constructed discourse portion of our methodology. We refer the reader to (Suri 1993) for a full description of how the methodology was used to extend a particular focusing framework to handle sentences of the form "SX because SY," although we summarize our technique and findings here.

Semantically-Slanted Discourse (SSD) Methodology

In previous literature on local focusing (e.g., (Sidner 1979), (Grosz, Joshi, & Weinstein 1983), (Brennan, Friedman, & Pollard 1987)), researchers used a small number of constructed texts to justify aspects of their focusing frameworks and to assess and compare focusing frameworks. However, they did not explicitly address how one should construct sets of texts in order to draw accurate conclusions about local focusing. The first part of our methodology is intended to help the researcher construct sets of texts (i.e., minimal pairs or minimal quadruples) that allow components of a focusing framework to be systematically isolated and thus allow one to appropriately assess focusing frameworks.

To appreciate the reasoning behind the first part of our methodology, or what we call our *Semantically-Slanted Discourse (SSD) Methodology*, recall that local focusing frameworks (including centering) are intended to capture the preferences for pronoun resolution *independent of* semantics, world knowledge, rhetorical relations and other kinds of pragmatics. Thus, they are intended to capture how one would resolve pronouns and update focusing information in discourse that is neutral in terms of these factors. Presumably in such texts, *only* focusing factors would affect pronoun resolution. In a semantically-non-neutral discourse, semantic fac-

tors (semantics, world knowledge, etc.) can override the preferences of the focusing framework by rejecting potential referents proposed by a focusing framework.

Taking this into account, in order to determine how best to process a particular type of complex sentence, we decided to construct discourses that have two important properties:

1. The set of discourses must be systematically constructed to ensure that each possible combination of focusing factors is represented in a discourse. We show how this is done below for a particular focusing framework and a single type of complex sentence.
2. The discourses are intentionally loaded or slanted for pronoun interpretation based on semantic factors. We call such a discourse *semantically-slanted* because the interpretation of all of the pronouns is fully determined by the semantic factors alone.

We contend that in a semantically-slanted discourse, if the text seems ambiguous or awkward, or if one needs to re-interpret a pronoun, then the focusing preferences for pronoun resolution are at odds with the preferences based on semantics, world knowledge, or other pragmatic factors. On the other hand, if the text seems acceptable/natural, then we contend the preferences for pronoun resolution based on focusing agree with preferences based on semantic slanting. Thus, gathering acceptability judgments about these systematically constructed semantically-slanted discourses should help us identify what the focusing preferences are, and thus how a focusing framework should be extended to handle a given type of complex sentence. This is the idea at the heart of our methodology.

Isolating the Complexity

In using semantically-slanted discourses to uncover an extension of a focusing algorithm, we must first make sure that the discourses we construct isolate the complexity under study. This factor influences the overall form of the discourses being constructed.

In order to isolate the complexity under study, we construct discourses of the following form, for which the interpretation of the NPs is fully determined by the semantics of the text and world knowledge:

Example 1

S1) Simple sentence
S2) Sentence with one level of complexity (i.e., having two clauses), introduced by the syntactic form of interest.
S3) Simple sentence

In examining linguistic judgments about such texts, our goal is to identify preferences imposed by the syntactic form of S2 for:

- resolving pronouns in S2
- updating the focusing data structures after S2 so that the pronouns of S3 can be correctly resolved in a manner that is consistent with resolving pronouns in a sentence following a simple sentence or another kind of complex sentence.

The motivation for having S1 be a simple sentence is to avoid any effect a complex sentence might have on the focusing data structures going into S2. Similarly, the motivation for having S3 be a simple sentence is to avoid any effect that a complex sentence structure in S3 might have on pronoun resolution in S3.

Systematic Construction of Discourses for "SX because SY" Sentences

Our methodology calls for the systematic construction of a set of discourses that ensures that all possible combinations of the focusing factors are represented. This will allow an extension of the focusing framework to emerge (since the discourses essentially capture all possibilities for an extension). To see the types of discourses our methodology calls for constructing, let us consider what would be needed to extend a particular focusing framework, RAFT/RAPR (described in (Suri 1993)), to handle resolving subject pronouns in sentences of the form "SX because SY" where SX and SY are simple sentences, and in a sentence following that type of sentence.

To address this question, we examined discourses that are "variations" of the form shown in Example 2.

Example 2

S1) Dodge was nearly robbed by an ex-convict the other night.
S2) [Dodge] captured [the ex-con] because [the ex-con] was so stupid and clumsy.
S3) Then [Dodge] called the police.

We needed to construct variations of this text in order to tease out how the various focusing factors interact.

The RAFT/RAPR algorithm prefers to resolve a subject pronoun (in a simple sentence) so that it co-refers with the subject of the previous sentence if the previous sentence is a simple sentence. The subject of a simple sentence is selected as (the contents of a focusing data structure that we call) the *Subject Focus* of that sentence. If this suggested referent (i.e., the previous sentence's Subject Focus) is rejected by inferencing with world knowledge, semantics, and pragmatics, then other elements in the previous discourse are tried in a specified order; this ordering is (indirectly) influenced by such things as whether a pronoun was used in the previous sentence (since pronouns are indicative of focus and therefore influence the focus computation for the previous sentence). Note, the algorithm prefers constancy in Subject Focus over shifting the Subject Focus. Some questions that must be answered in coming up with an extension of RAFT/RAPR in handling "SX because SY" sentences are:

1. How should Subject(SY) be resolved? I.e., should the algorithm prefer that it co-refers with Subject(S1) or Subject(SX)?
2. How should Subject(S3) be resolved? I.e.,
   - Preferring Subject(SX) always?
   - Preferring Subject(SY) always?
   - Preferring Subject(SX) or Subject(SY) depending on which is pronominalized?
   - Preferring Subject(SX) or Subject(SY) depending on which is co-referential with Subject(S1)?
   - Based on some other preference?

The answers to these (and all such similar questions) constitute a decision about how and whether the complex sentence should be segmented, and how to weigh the influences of the various focusing factors such as pronominalization and focus history in resolving pronouns and in updating the focusing data structures.

To see how we make up the variations to cover all possibilities, consider this abstract view of the text which indicates the NPs we are interested in:

Example 3

S1) Subject(S1) ... Direct-Object(S1)
S2) Subject(SX) ... Direct-Object(SX)
because Subject(SY) ...
S3) Subject(S3) ...

To find an extension of the focusing algorithm, we need to make up variations of the text which capture all the different ways the grammatical roles, focus history, and pronominalization (i.e., the focusing factors) might interact in determining the referent of the pronoun in Subject(S3). We make up text variations corresponding to variations of the following parameters:

1. Whether Subject(S1) is the ex-convict or Dodge. ("An ex-convict nearly robbed Dodge the other night" vs. "Dodge was nearly robbed by an ex-convict the other night.")
   - Note that the Direct-Object(S1) will always introduce the other actor. This helps test the focusing factor of grammatical role.
2. Whether Subject(SX) of S2 is the ex-convict or Dodge. ("[Dodge] captured [the ex-convict] because [the ex-convict] was so stupid and clumsy" vs. "[The ex-convict] woke [Dodge] up because [the ex-convict] was so stupid and clumsy.")
   - Notice that 1 and 2 together will vary whether or not the Subject(S1)=Subject(SX).
   - Note if Subject(S1)≠Subject(SX) then Direct-Object(S1)=Subject(SX), again testing grammatical role effects.
3. Whether Subject(SY) of S2 is the ex-convict or Dodge. ("[The ex-convict] tied [Dodge] up because [the ex-convict] didn't want any trouble" vs. "[The ex-convict] tied [Dodge] up because [Dodge] wasn't co-operating.")
   - 1 and 3 taken together alter whether or not Subject(S1)=Subject(SY).
   - 2 and 3 taken together alter whether or not Subject(SX)=Subject(SY).
   - Because the focusing algorithm prefers constancy in Subject Focus history, these alternations help us decide whether S1 or SX is more important in resolving pronouns in SY.
4. Whether Subject(S3) is the ex-convict or Dodge.[5] ("Then [the ex-convict] was arrested by the police" vs. "Then [Dodge] started screaming for help.")
   - This parameter (in conjunction with the others) determines whether Subject(S3)=Subject(SX) or Subject(S3)=Subject(SY) or neither (i.e., Subject(S3)=Direct-Object(SY)).
   - Similar to 3 above, this parameter helps us determine how the NPs in S2 affect the resolution of the pronoun in S3.
5. Whether Subject(SX) was pronominalized.
6. Whether DirectObject(SX) was pronominalized.
7. Whether Subject(SY) was pronominalized.
   - 5-7 help check how various patterns of pronominalization might affect the processing.

By generating texts for all combinations of different values of these parameters, we are able to control for the influence of each focusing factor. Essentially, taken together, the texts capture all different grammatical roles, focus history patterns, and patterns of pronominalization.

The result of this procedure is a set of texts which can be presented to native speakers for judgments. The idea is that in texts which are judged to be acceptable, the reader sees no conflicts, so the focusing factors (grammatical role, pronoun history, and focus history) must agree with the semantic slanting. In texts which are judged unacceptable or ambiguous, there must be a conflict between these two sets of factors. Thus, gathering such judgments should identify an extension of the focusing algorithm. On the basis of the resulting judgments, a possible extension of the focusing algorithm can be identified.

Finding an Extension Based on Judgments

To see how this methodology can be used to find an extension for a focusing framework, let us concentrate on determining how RAFT/RAPR should compute the focusing data structures for an "SX because SY" sentence in order to correctly resolve a subject pronoun in a simple sentence following an "SX because SY" sentence. In considering this question, recall that

[5] Recall that when Subject(S3) is pronominalized, the referent of Subject(S3) is determined by the semantic slanting of the text.

RAFT/RAPR prefers to resolve a subject pronoun in a simple sentence with the Subject Focus of the previous sentence. Thus, the specific question we need to address is how to compute the Subject Focus of an "SX because SY" sentence.[6]

Again, the factors that might determine how to compute the subject focus of a sentence are:

1. syntactic form: for example, perhaps the Subject Focus should be computed as Subject(SX) (or, alternately, Subject(SY)) (regardless of other factors). This would be appropriate if readers prefer to resolve the subject of the subsequent sentence so that it co-specifies Subject(SX) (or, alternately Subject(SY)), regardless of other focusing factors.
2. pronominalization: for example, perhaps the Subject Focus should be computed as Subject(SX) or as Subject(SY) depending on which is pronominalized, regardless of other factors (unless both or neither are pronominalized). This would be appropriate if readers prefer to resolve the subject of the subsequent sentence on the basis of whether Subject(SX) or the Subject(SY) was pronominalized.
3. focus history: for example, perhaps Subject Focus should be computed as Subject(SX) or Subject(SY) depending on which co-specified the Subject Focus of the sentence preceding the "SX because SY" sentence.
4. Perhaps the interaction of two or more of the above factors (syntactic form, pronominalization and focusing history) might influence how the Subject Focus should be computed.

Our findings indicate that the syntactic form alone seems to most greatly influence what should be chosen as the Subject Focus of an "SX because SY" sentence. In particular, we found that readers prefer to resolve the subject of a sentence following an "SX because SY" sentence so that the subject co-specifies Subject(SX). This indicates that RAFT/RAPR should compute the Subject Focus as Subject(SX). Consider how the following two judged discourses support this conclusion.

Example 4

(S1) Dodge was robbed by an ex-convict the other night.
(S2) The ex-convict tied him up because he wasn't cooperating.
(S3) Then he took all the money and ran.

Notice that in the discourse in Example 4 the semantic slanting should lead to the interpretation of S3 as "Then [the ex-conflict] took all the money and ran." Thus, in this text, the semantic slanting favors resolving Subject(S3) as Subject(SX), and thus it would favor computing the Subject Focus of S2 to be Subject(SX). However, notice that the pronominalization in the text favors computing the Subject Focus of S2 (which will be used to resolve Subject(S3)) to be Subject(SY) since Subject(SY) is pronominalized in S2, but Subject(SX) is not. The focus history would also favor the reading in which Subject(S3) is Dodge (i.e., co-refers with the Subject(SY)) since Dodge was the Subject Focus of S1 and Dodge occurs in S2 as a subject.

Even though only the factor of syntactic form favors the interpretation that agrees with semantic slanting, of the 33 subjects in our experiment, 91% judged this discourse as acceptable. This supports the hypothesis that the syntax is the most important focusing factor and that it favors resolving a subject in a simple sentence following an "SX because SY" sentence so that the subject co-specifies Subject(SX).

The above hypothesis is further supported by the judgments given on a second discourse:

Example 5

(S1) Dodge was robbed by an ex-convict the other night.
(S2) The ex-convict tied him up because he wasn't cooperating.
(S3) # Then he started screaming for help.

In Example 5, one would expect subjects to judge the discourse to be "awkward" or "ambiguous" if the syntax factor overrides other focusing factors and causes Subject(SX) to be the preferred referent for the subject of the subsequent sentence. This is because the semantic slanting should lead to the interpretation of S3 as "Then [Dodge] started screaming for help," i.e., an interpretation for which Subject(S3)≠Subject(SX). In fact, most subjects (90%), did judge the discourse as awkward or ambiguous.[7]

To reiterate, in Example 5:

- On the basis of on semantic slanting, Subject(S3)=Subject(SY) and Subject(S3)≠Subject(SX).
- Subject Focus history would favor Subject(SY) for the Subject Focus of S2 since Subject(SY)=Subject(S1) (i.e., the Subject Focus of S1).
- Pronominalization would favor Subject(SY) for the Subject Focus of S2 since the referent of Subject(SY) is pronominalized in S2, but Subject(SX) is not.
- The discourse was judged awkward or ambiguous.

In Example 4, the focus history and pronominalization favor computing the SF(S2) to be Subject(SY), the semantic slanting indicates Subject(S3) is Subject(SX), and the text was judged acceptable.

Taken together, these judgments suggest that the reader prefers to resolve Subject(S3) with Subject(SX) regardless of the other focusing factors since this would explain the judgments as follows: In Example 5, the interpretation of Subject(S3) indicated by this focusing preference would be at odds with the interpretation forced by the semantic slanting and the text was

[6]For other focusing frameworks, we would be concerned with a different specific question.

[7]As a result we have labeled (S3) with a "#" which traditionally denotes pragmatic ill-formedness.

judged awkward; in Example 4, the interpretation indicated by this focusing preference would agree with the interpretation forced by the semantic slanting and the text was judged acceptable.

Therefore, we conclude that RAFT/RAPR should compute the Subject Focus of an "SX because SY" sentence to be Subject(SX).

Conclusions

The notion of local focusing and its influence on pronoun resolution has been found useful in many aspects of NLP. However, previous work on local focusing has ignored complex sentences even though they are prevalent in naturally-occurring text. The problem that we faced was one of determining a reasonable way to extend a focusing algorithm to handle these sentences. Previous methodology (i.e., using semantically neutral text) was too simplistic and nearly impossible to utilize. A solely corpus-based analysis is impossible because of the variety of *a priori* decisions that needed to be made and because of the complexity of interaction among factors in naturally-occurring discourses. This work presents a methodology that calls for the systematic construction of texts. It relies on the potential tension of semantic factors with focusing factors to identify possible extensions of a focusing framework to account for a particular kind of complex sentence. The methodology has been used to extend a focusing framework to handle one type of complex sentence.

Furthermore, as explained in detail in (Suri 1993), this methodology can also be used to *compare* local focusing frameworks. Thus, this methodology allows one to study focusing phenomena and algorithms related to focusing phenomena.

References

Brennan, S. E.; Friedman, M. W.; and Pollard, C. J. 1987. A centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, 155–162.

Brennan, S. E. 1993. Reference and local structure in interactive discourse. In the proceedings of the Workshop on Centering Theory in Naturally-Occurring Discourse, Institute for Research in Cognitive Science, University of Pennsylvania.

Carter, D. 1987. *Interpreting Anaphors in Natural Language Texts.* New York: John Wiley and Sons.

Dahl, D., and Ball, C. N. 1990. Reference resolution in PUNDIT. Technical Report CAIT-SLS-9004, UNISYS, Paoli, PA.

Dahl, D. A. 1986. Focusing and reference resolution in PUNDIT. In *Proceedings of the 1986 National Conference on Artificial Intelligence*, 1083–1088.

Grosz, B. J.; Joshi, A. K.; and Weinstein, S. 1983. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, 44–50.

Hoffman, B., and Turan, Ü. D. 1993. Word order and centering in Turkish. In the proceedings of the Workshop on Centering Theory in Naturally-Occurring Discourse, Institute for Research in Cognitive Science, University of Pennsylvania.

Kameyama, M.; Passonneau, R.; and Poesio, M. 1993. Temporal centering. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, 70–77.

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

Kameyama, M. 1993. Intrasentential centering. In the proceedings of the Workshop on Centering Theory in Naturally-Occurring Discourse, Institute for Research in Cognitive Science, University of Pennsylvania.

Linson, B. 1993. *A Distributional Analysis of Discourse Transitions.* In the proceedings of the Workshop on Centering Theory in Naturally-Occurring Discourse, Institute for Research in Cognitive Science, University of Pennsylvania.

Palmer, M. S.; Dahl, D. A.; Schiffman, R. J.; Hirschman, L.; Linebarger, M.; and Dowding, J. 1986. Recovering implicit information. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, 10–19.

Sidner, C. L. 1979. *Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse.* Ph.D. Dissertation, MIT.

Suri, L. Z. 1993. *Extending Focusing Frameworks to Process Complex Sentences and to Correct the Written English of Proficient Signers of American Sign Language.* Ph.D. Dissertation, University of Delaware. Available as Dept. of CIS Technicial Report TR-94-21.

Walker, M.; Iida, M.; and Cote, S. 1992. Japanese discourse and the process of centering. Technical Report 92-14, The Institute for Research in Cognitive Science, The University of Pennsylvania.

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

Walker, M. 1993. *Initial Contexts and Shifting Centers.* In the proceedings of the Workshop on Centering Theory in Naturally-Occurring Discourse, Institute for Research in Cognitive Science, University of Pennsylvania.

Acknowledgments

We thank Jeff Lidz and John Hughes for comments on this research. We thank our informants for their help and time in providing us with judgments.