

Pronominalization in Generated Discourse and Dialogue

Charles B. Callaway

Istituto per la Ricerca Scientifica e
Tecnologica (ITC-irst), Italy
callaway@irst.itc.it

James C. Lester

Department of Computer Science
North Carolina State University, USA
lester@adm.csc.ncsu.edu

Abstract

Previous approaches to pronominalization have largely been theoretical rather than applied in nature. Frequently, such methods are based on Centering Theory, which deals with the resolution of anaphoric pronouns. But it is not clear that complex theoretical mechanisms, while having satisfying explanatory power, are necessary for the actual generation of pronouns. We first illustrate examples of pronouns from various domains, describe a simple method for generating pronouns in an implemented multi-page generation system, and present an evaluation of its performance.

1 Introduction

Pronominalization is an important element in the automatic creation of multi-paragraph and multi-page texts using natural language generation (NLG). Authors routinely use pronouns in texts spanning all types of genres, such as newspaper copy, science fiction and even academic papers. Indeed, without pronouns, texts quickly become confusing as readers begin to pay more attention to the writing style than to the content that makes the text informative or enjoyable (Callaway and Lester, 2001a). Even worse, incorrect pronouns can lead readers to misinterpret the text or draw unsound inferences.

Furthermore, current pronominalization strategies are ill-equipped to deal with the wide variety of reasons that pronouns are used in naturally occurring texts. Almost without exception, they focus on

anaphoric pronouns as described in Focus/Centering Theory (Webber, 1979; Sidner, 1983; Grosz and Sidner, 1986; Walker, 1998), ignoring the multitude of other possible types. However, it is certainly true that authors make use of pronouns which are not motivated by anaphoric reference.

In addition, because such approaches are oriented towards anaphora resolution during parsing, they ignore structures such as the discourse plan which are present during generation but not parsing. A typical discourse plan can include vital information for pronominalization such as time and clause boundaries, ordering of propositions, and semantic details verbal arguments. Current approaches based on Centering algorithms thus attempt to recreate a text coherence structure that duplicates work already done by the discourse planner.

Finally, there are significant obstacles to verifying the correctness of existing pronominalization algorithms for any pronominalization theory (Not, 1996; Yeh and Mellish, 1997; McCoy and Strube, 1999; Henschel et al., 2000; Kibble and Power, 2000): the lack of natural language generation systems that can produce large enough texts to bring discourse-level processes into play. Because of this, researchers are forced to simulate by hand how their algorithms will work on a given text. It is also not sufficient to use template generation systems to perform this task because they lack the low-level discourse representation needed to provide the information upon which most algorithms base their decisions.

In this paper we first summarize related work in both anaphora resolution and anaphora generation. We next describe the range of pronoun types

that we found in a wide variety of texts. We proceed to describe an algorithm for determining appropriate pronominalizations that uses existing NLG structures and simple numeric techniques. We also briefly describe an implemented generation system that contains enough low-level discourse information to motivate pronominalization decisions using this method. Finally, we quantitatively demonstrate the performance of this simple numerical approach in both a newspaper and fictional narrative domain.

2 Background and Related Work

Because most NLG systems have focused on linguistic phenomena at the paragraph level and below, there has been intensive investigation into the core areas of generation that are required to produce them: discourse planning, sentence planning and surface realization. Since pronouns are more likely to be a multiparagraph, discourse-level phenomenon, it has been possible to ignore their inclusion into working NLG systems which are not called upon to generate lengthy passages.

Indeed, most work on pronouns in computational linguistics has come under the heading of anaphora resolution as an element of parsing rather than the heading of pronominalization as an element of generation. Since discourse anaphora resolution was first studied theoretically (Grosz, 1977; Webber, 1979; Sidner, 1983; Grosz and Sidner, 1986), it has come to be dominated by Centering Theory (Grosz et al., 1995; Di Eugenio, 1998; Walker, 1998) which proposes rules for the determination of focus and salience within a given segment of discourse. Relatively little work has been done on alternate approaches to pronoun resolution (Hobbs, 1976; Baldwin, 1995).

While many NLG researchers have attempted to transfer the ideas of Centering Theory to generation (Not, 1996; Yeh and Mellish, 1997; McCoy and Strube, 1999; Henschel et al., 2000; Kibble and Power, 2000), there has yet been no substantial return contribution to the field of anaphora resolution. There are two principal reasons for this. First, it is extremely difficult to create an NLG system that generates the large quantity of texts needed to exhibit discourse-level phenomena while consistently employing the deep linguistic representations

needed to determine appropriate pronominal forms. Second, Centering Theory is still vague on the exact definition of terms such as “segment” (Poesio et al., 1999a), making it difficult to create a mutually agreeable implementation.

An additional area of NLG research that deals with pronouns is that of referring expression generation (Appelt, 1985; Heeman and Hirst, 1986; Claassen, 1992; Dale, 1992), which attempts to find the optimal noun phrase (whether full description, definite description, deixis, pronoun, or reduced noun phrase) to enable a reader to mentally select the intended referent from the set of possible referents (Reiter and Dale, 1997). Comparatively, referring expression generation is a process for local disambiguation and is not generally concerned with single phenomena spanning multiple paragraphs. Because of this, and because the domains and genres we have studied typically do not involve sets of very similar referents, we concentrate on discourse-motivated sources of pronominalization.

3 Examples of Pronominalization

Pronominalization is the appropriate determination, marking and grammatical agreement of pronouns (*he, she, their, herself, it, mine, those, each other, one, etc.*) as a short-hand reference to an entity or event mentioned in the discourse. As with anaphora resolution, the task of a pronominalization algorithm is to correctly predict which pronoun a person would prefer in the same situation. The range of possibilities includes leaving the noun phrase as it is, reducing it by removing some of its modifiers, or replacing it with a pronoun construction.

Our corpora analyses have identified a number of motivations for converting nouns into pronouns:

1. *Anaphoric pronouns*: These are the most-studied cases of pronoun occurrences, which sequentially follow a specific entity known as the *referent*. Anaphors are divided into two classes, *short-distance* (within the same sentence) and *long-distance* (previous sentences).

But John_i had never been to New Orleans, and he_i couldn't remember if anyone in his_i family had either.

2. *Cataphoric pronouns*: According to Quirk *et*

al. (1985), cataphors are those pronouns which occur before their referents in the linear flow of text within the same sentence, where the pronoun is either at a lower structural level or is part of a fronted circumstantial clause or prepositional phrase which could have appeared after the reference. Additionally, this category could include clefting pronouns.

Before he_i joined the navy, Gerald $_i$ made peace with his family.

3. *Pronouns Lacking Textual Antecedents*: This category includes document deixis (via a demonstrative pronoun), authorial or reader reference, and situational pronouns.

This is the first document to show . . .

We discuss these strategies in the next section.

The group had never seen anything like *it*.

4. *Reflexive and Reciprocal Pronouns*: Most verbs use special pronouns when the subject and object corefer. A discourse history algorithm can employ that knowledge to mark reflexive and reciprocal pronouns appropriately.

Kittens $_i$ often watch *themselves $_i$* in mirrors.

Baby lions $_j$ tackle *each other $_j$* when playing.

5. *Partitive pronouns*: It is important to know conceptually what it is that the pronoun is trying to replace. Otherwise, it becomes impossible to achieve the types of pronominalizations that authors are routinely capable of creating. This requires accurate information in the knowledge base or linguistic structure from which the sentences are derived.

As the horses ran by, she roped *one*.

* As the horses ran by, she roped *it*.

* As the horses ran by, she roped *them*.

In addition to these motivations, we identified several factors that prevent pronouns from occurring where they otherwise might:

6. *Pronouns across boundaries*: After a chapter, section or other obvious boundary, such as a change in time, place, or both, as in (McCoy and Strube, 1999), authors will typically “reset” pronominalization just as if it were the

beginning of the entire text. Antecedent references that break these boundaries are sometimes marked by the authors in academic texts:

As we saw in the previous section, . . .

7. *Restrictions from modifiers*: Because pronouns cannot have modifiers like nouns, adding an adjective, relative clause, or some other modifier prevents a noun from being replaced by a pronoun. For instance:

The mayor had already read the *full proposal*.

* The mayor had already read the *full it*.

8. *Focused nouns*: Especially after a vocally stressed discourse marker (Wolters and Byron, 2000) or some other marked shift in topic, a word that normally would be pronominalized is often not, as in this example:

. . . and you frequently find that mice occupy an important part of the modern medical laboratory. In other words, *mice* are especially necessary for diagnosing human cancers . . .

9. *Semantic and syntactic considerations*: A small number of semantic relations and syntactic constructions prohibit pronominalization:

* The stranger was just called *him*. (Bob)

* Roberta was no longer a *her*. (child)

* The father, a tyrant of a *him*, . . . (man)

10. *Optional pronominalization*: Often there are borderline cases where some authors will use pronouns while others won't. A single algorithm may be tuned to match a particular author's style, but parameterization will be necessary to match a variety of styles. Thus it is extremely difficult to exactly match any particular text without having the ability to adjust the pronominalization algorithm.

Pronominalization occurs equally as often in exposition as in dialogue, but dialogue can have slightly different pronominalizations depending on the relationship between the utterer and the hearer:

11. *Speaker self-reference*:

“John thinks John will go find John's shoes,”
John said.

changes to first person singular pronouns:

“I think I will go find my shoes,” John said.

12. *Speaker references hearer(s)*:

“Mary should go find Mary’s shoes,” John said.

changes to second person pronouns:

“You should go find your shoes,” John said.

13. *Reference to speaker and hearer (or to speaker and a third party)*:

“John and Mary should go find John and Mary’s shoes,” John said.

changes to first person plural pronouns:

“We should go find our shoes,” John said.

14. *Reference to a third party*:

“Bob and Mary went to eat Bob and Mary’s breakfast,” John said.

changes to third person plural pronouns:

“They went to eat their breakfast,” John said.

15. Finally, the treatment of pronouns differs depending if they are inside or outside of the direct quotation. For example:

“Oh man, I forgot to eat my breakfast!” John muttered to himself while grabbing his shoes.

Although this enumeration is surely incomplete, it provides a basic description of the types of phenomena that must be handled by a generation system in order to produce text with the types of pronouns found in routine human-produced prose.

4 Architectural Concerns

In order to correctly account for these phenomena during generation, it is necessary to have detailed information about the underlying discourse structure. Although a template generation system could be augmented to record this information, in practice only deep structure, full-scale NLG systems have the requisite flexibility. Because a pronominalization algorithm typically follows the discourse planner, it frequently has access to the full discourse plan.

A typical discourse plan is a tree structure, where internal nodes represent structuring relations while

leaf nodes represent individual sentential elements that are organized semantically. In addition, the elements of the discourse tree are typically rooted in the semantic knowledge base which the discourse planner drew from when constructing the discourse plan. The discourse plan supplies the following information that is useful for pronominalization:

- *Linearization*: The sequencing information stored in the discourse tree can be used to motivate anaphoric and cataphoric pronouns as shown in items 1 & 2 of Section 3.
- *Semantic Structure*: The original subgraphs (or semantic subnetworks) derived from the knowledge base can motivate content vs. situational knowledge (item 3) reflexive and reciprocal pronouns via argument lists (item 4), partitive pronouns (item 5), and the existence of NP modifiers (item 7), and can identify semantic types in relations (item 9).
- *Discourse Structure*: The rhetorical relations that hold between different sentences typically imply where section boundaries are located (item 6), indicate what types of discourse markers are employed (item 8), and in the case of dialogue, know which actors are speaking, listening, or not present (items 11-15).

This detailed knowledge of the discourse is available to an implemented pronominalization component utilizing any theory, including Centering theory. We thus now turn our attention to what role this information plays in a pronominalization algorithm.

5 A Simple Pronominalization Algorithm

At an abstract level, the pronominalization algorithms derived from Centering theory are easily expressed: if Centering theory predicts a pronoun would be used in anaphora resolution in a given segment of text, then generate the appropriate pronoun. While this works for many cases of anaphoric pronouns [84.7% in (McCoy and Strube, 1999), 87-90% in (Henschel et al., 2000)], we have seen that these form only a subset of the potential reasons for pronominalization. Furthermore, this approach assumes that the discourse tree was constructed with Centering theory in mind.

Given:

LNE , the linearized list of nominal elements
 NE , the current nominal element
 $SEEN$, the list of encountered nominal elements
 D , the dialogue state of the current leaf node
 RS , the rhetorical structure near the leaf node
 SC , the sentence counter

Do:

```

 $SEEN \leftarrow \phi, SC \leftarrow 0$ 
while  $LNE \neq \phi$  do
   $NE \leftarrow first(LNE)$ 
  if  $NE \notin SEEN$ 
    then  $reset-counters(NE)$ ,
           $SEEN \leftarrow SEEN \oplus NE$ 
  else  $update-counters(NE)$ 
   $D \leftarrow updateDialogueState()$ 
   $RS \leftarrow updateLocalRhetoricalStructure()$ 
  if  $(topic-shift \vee time-shift) \in RS$ 
    then  $SC \leftarrow SC + 10$ 
  else if  $modifiers(NE, RS) = \phi \wedge$ 
            $(special-relation \vee appositive) \notin RS$ 
    if  $D == QuotedDialogue$ 
      then  $mark(quoted-pronoun(NE, RS))$ 
    else if  $subject-matches-object(NE, RS)$ 
      then  $mark(ReflexivePronoun)$ 
    else if  $sent-distance(NE, SC) = 0$ 
      then  $mark(MultipleInSentencePronoun)$ 
    else if  $3 \leq sent-distance(NE, SC) < 1$ 
           and  $nominal-distance(NE) < 3$ 
      then  $mark(LongDistancePronoun)$ ,
    else if  $recency(NE) > 3$ 
      then  $mark(ShortDistancePronoun)$ ,
   $LNE \leftarrow remove-first(LNE), SC \leftarrow SC + 1$ 

```

Figure 1: The Pronominalization Algorithm

However, it is not clear that Centering theory itself is necessary in generation, let alone its accompanying algorithms and data structures. Because Centering theory is typically applied to parsing (which starts with no discourse tree), it may not be the most efficient technique to use in generation (which has a complete discourse tree available for inference).

Instead, we attempted to determine if the information already present in the discourse tree was enough to motivate a simpler algorithm based on the following available data:

- *Ordered sequence of nominal elements:* Because the discourse tree is linearized and individual leaves of the tree annotate which elements have certain semantic roles, a very good guess can be made as to which nominal elements precede others at the clause level.
- *Known paragraph and sentence boundaries:* Analysis of the rhetorical structure of the dis-

course tree allows for the determination of boundaries and thus the concept of metric distance between elements.

- *Rhetorical relations:* The rhetorical relations can tell us which nominal elements follow discourse markers and which are used reflexively or reciprocally.
- *Dialogue:* By recording the participants in dialogue, the discourse tree allows for the appropriate assignment of pronouns both inside and outside of the direct quote itself.

The algorithm we developed considers the current discourse leaf node and the rhetorical structure above it, and also makes use of the following data:

- *Nominal element distance:* How many total (non-distinct) nominal elements ago a particular element was last used.
- *Recency:* How many distinct nominal elements have been seen since its last use.
- *Sentential distance:* How many sentences (prototypical clauses) have appeared since the last usage of this nominal element.

The algorithm itself (Figure 1) is best characterized as a *counting method*, that is, it loops once through the linearized list of nominal elements and makes pronominalization decisions based on the local information described above, and then updates those numerical counters. Numerical parameters (e.g., $recency(NE) > 3$) are derived from empirical experimentation in generating multi-page prose in a narrative domain.

While it lacks the explanatory power of a relatively mature linguistic theory, it also lacks the accompanying complexity and is immediately applicable to real-world deep generation systems. The algorithm is traced in Figure 2, although due to space limitations some phenomena such as dialogue, long distance and reflexive pronouns are not shown.

6 Implementation and Evaluation

STORYBOOK (Callaway and Lester, 2001b; Callaway and Lester, in press) is an implemented narrative generation system that converts a pre-existing

<p>Sentences as seen by the reader (antecedents underlined, pronouns in bold):</p> <p>Now, it happened that a <u>wolf</u>₁, a very cruel, greedy <u>creature</u>₂ also heard <u>Little Red Riding Hood</u>₃ as she₄ passed, and he₅ longed to eat her₆ for his₇ <u>breakfast</u>₈. But he₉ knew <u>Hugh</u>₁₀, the <u>woodman</u>₁₁, was at <u>work</u>₁₂ very near with his₁₃ great <u>dog</u>₁₄.</p>
<p>Sentences as produced by the discourse planner before revision:</p> <p>S1: Now, it happened that a <u>wolf</u>₁, a very cruel, greedy <u>creature</u>₂ also heard <u>Little Red Riding Hood</u>₃ as <u>Little Red Riding Hood</u>₄ passed.</p> <p>S2: The <u>wolf</u>₅ longed to eat <u>Little Red Riding Hood</u>₆ for the <u>wolf's</u>₇ <u>breakfast</u>₈.</p> <p>S3: But the <u>wolf</u>₉ knew <u>Hugh</u>₁₀, the <u>woodman</u>₁₁, was at <u>work</u>₁₂ very near with <u>Hugh's</u>₁₃ great <u>dog</u>₁₄.</p>
<p>Each noun element is processed in the order linearized from the discourse plan:</p> <ol style="list-style-type: none"> 1. The first mention of <i>wolf</i>₁ in the narrative resets its discourse history entry. 2. <i>Creature</i>₂ is the second mention of <i>wolf</i>, but it is in an appositive structure (see pronoun category #9). 3. <i>LRRH</i>₃ was mentioned just before in the prior paragraph, but “Now,” is a prosodic discourse marker (see pronoun category #8), thus <i>modifiers</i>(<i>NE</i>, <i>RS</i>) ≠ ϕ. 4. For <i>LRRH</i>₃ and <i>LRRH</i>₄, <i>sentence-distance</i>(<i>NE</i>, <i>SC</i>) = 0 resulting in a <i>multiple-in-sentence-pronoun</i>. 5. <i>Sentence-distance</i>(<i>NE</i>, <i>SC</i>) = 1, but <i>recency</i>(<i>NE</i>) = 2, resulting in a <i>short-distance-pronoun</i>. 6. Similarly, <i>LRRH</i>₆ is converted into a <i>short-distance-pronoun</i>. 7. As with element #4, this is a case resulting in a <i>multiple-in-sentence-pronoun</i>. 9. As with element #5, this is a case resulting in a <i>short-distance-pronoun</i>. 10. The first mention of <i>Hugh</i>₁₀ in the narrative resets its discourse history entry. 11. As with element #2, the discourse plan reports that this is an appositive. 13. Finally, <i>Hugh</i>₁₃ is repeated in the same sentence.

Figure 2: A Brief Trace of the Pronominalization Algorithm for Anaphoric Pronouns from STORYBOOK

narrative (discourse) plan into a multi-page fictional narrative in the fairy tale domain. Using a pipelined generation architecture, STORYBOOK performs pronominalization before sentence planning, and includes a revision component that is sensitive to pronominalization choices during clause aggregation. A previous large-scale evaluation of STORYBOOK (Callaway and Lester, 2001a) which included both a full version and a version with the pronominalization component ablated showed that including such a component significantly increases the quality of the resulting prose.

However, there are significant practical obstacles to comparing the performance of different pronominalization algorithms using corpus matching criteria instead of “quality” as evaluated by human judges. Because systems that can handle a large quantity of text are very recent and because it can require years to create and organize the necessary knowledge to produce even one multi-paragraph text, much research on anaphora generation has instead relied on

one of two techniques:

- *Checking algorithms by hand*: One verification method is to manually examine a text, identifying candidates for pronominalization and simulating the rules of a particular theory. However, this method is prone to human error.
- *Checking algorithms semiautomatically*: Other researchers opt instead to annotate a corpus for pronominalization and their antecedents as well as the pronoun forms that should occur, and then simulate a pronominalization algorithm on the marked-up text (Henschel et al., 2000). Similarly, this approach can suffer from interannotator agreement errors (Poesio et al., 1999b).

To verify our pronominalization algorithm more rigorously, we instead used the STORYBOOK deep generation system to recreate pre-existing multi-page texts with automatically selected pronouns.

	McCoy & Strube NYT News	Henschel et al. NYT News	STORYBOOK NYT News	STORYBOOK LRRH Narrative
Animate Anaphora	370/437 (84.7%)	N/A	415/449 (92.4%)	170/174 (97.7%)
All Anaphora	N/A	469/527 (89.0%)	441/475 (92.8%)	177/181 (97.8%)
Cataphora	N/A	N/A	1/2 (50.0%)	1/2 (50.0%)
Dialogue	N/A	N/A	46/46 (100.0%)	65/65 (100.0%)
Deixis	N/A	N/A	9/9 (100.0%)	None present
Reflex./Recip.	N/A	N/A	5/6 (83.3%)	2/2 (100.0%)
Partitive	N/A	N/A	1/2 (50.0%)	1/1 (100.0%)

Table 1: Pronouns Correct by Algorithm/Text vs. Pronoun Type

Without a full-scale implementation, it is impossible to determine whether an algorithm performs imperfectly due to human error, a lack of available corpus data for making decisions, or if it is a fault with the algorithm itself.

Using the algorithm described in Figure 1, we modified STORYBOOK to substitute the types of pronouns described in Section 3. We then created the discourse plan and lexicon necessary to generate the same three articles from the New York Times as (McCoy and Strube, 1999). The results for both the newspaper texts and the Little Red Riding Hood narrative described in (Callaway and Lester, in press) are shown in Table 1.

With the same three texts from the New York Times, STORYBOOK performed better than the previous reported results of 85-90% described in (McCoy and Strube, 1999; Henschel et al., 2000) on both animate and all anaphora using a corpus matching technique. Furthermore, this was obtained solely by adjusting the *recency* parameter to 4 (it was 3 in our narrative domain), and without considering other enhancements such as gender/number constraints or domain-specific alterations.¹

7 Conclusions

Pronominalization is an important element in the automatic creation of multi-paragraph and multi-page

¹It is important to note, however, that our counts of pronouns and antecedents do not match theirs. This may stem from a variety of factors, such as including single instances of nominal descriptions, whether dialogue pronouns were considered, and if borderline quantifiers and words like “everyone” were counted. The generation community to-date has not settled on standard, marked corpora for comparison purposes as has the rest of the computational linguistics community.

texts. Previous approaches, based largely on theoretical approaches such as Centering Theory, deal exclusively with anaphoric pronouns and have complex processing and definitional requirements.

Given the full rhetorical structure available to an implemented generation system, we devised a simpler method of determining appropriate pronominalizations which was more accurate than existing methods simulated by hand or performed semi-automatically. This shows that approaches designed for use with anaphora resolution, which must build up discourse knowledge from scratch, may not be the most desirable method for use in NLG, where discourse knowledge already exists. The positive results from our simple counting algorithm, after only minor changes in parameters from a narrative domain to that of newspaper text, indicates that future high-quality prose generation systems are very near.

8 Acknowledgements

We would like to thank Michael Young and Renate Henschel for their helpful comments; Kathy McCoy very quickly provided the original 3 NYT articles upon request; the anonymous reviewers whose comments greatly improved this paper. Support for this work was provided by ITC-irst and the IntelliMedia Initiative of North Carolina State University.

References

- Douglas E. Appelt. 1985. Planning English referring expressions. *Artificial Intelligence*, 26:1–33.
- Frederick Baldwin. 1995. *CogNIAC: A Discourse Processing Engine*. Ph.D. thesis, The University of Pennsylvania, Philadelphia, PA.

- Charles B. Callaway and James C. Lester. 2001a. Evaluating the effects of natural language generation on reader satisfaction. In *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*, pages 164–169, Edinburgh, UK.
- Charles B. Callaway and James C. Lester. 2001b. Narrative prose generation. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, pages 1241–1248, Seattle, WA.
- Charles B. Callaway and James C. Lester. 2003. Narrative prose generation. *Artificial Intelligence*. In press.
- Wim Claassen. 1992. Generating referring expressions in a multimodal environment. In R. Dale, E. Hovy, D. Rosner, and O. Stock, editors, *Aspects of Automated Natural Language Generation*, pages 247–62. Springer-Verlag, Berlin.
- Robert Dale. 1992. *Generating Referring Expressions*. MIT Press.
- Barbara Di Eugenio. 1998. Centering in Italian. In Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors, *Centering in Discourse*. Oxford University Press, Cambridge, MA.
- Barbara J. Grosz and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2).
- Barbara J. Grosz. 1977. The representation and use of focus in a system for understanding dialogs. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pages 67–76, Cambridge, MA.
- Peter Heeman and Graeme Hirst. 1986. Collaborating on referring expressions. *Computational Linguistics*, 12(3):351–382.
- Renate Henschel, Hua Cheng, and Massimo Poesio. 2000. Pronominalization revisited. In *COLING–2000: Proceedings of the 18th International Conference on Computational Linguistics*, Saarbruecken, Germany.
- Jerry R. Hobbs. 1976. Pronoun resolution. Technical Report 76-1, Department of Computer Science, City College, CUNY, New York, NY.
- Roger Kibble and Richard Power. 2000. An integrated framework for text planning and pronominalization. In *Proceedings of the First International Conference on Natural Language Generation*, pages 194–200, Mitzpe Ramon, Israel.
- Kathleen F. McCoy and Michael Strube. 1999. Taking time to structure discourse: Pronoun generation beyond accessibility. In *Proceedings of the Twenty-First Conference of the Cognitive Science Society*, pages 378–383, Vancouver, CA, August.
- Elena Not. 1996. A computational model for generating referring expressions in a multilingual application domain. In *COLING–1996: Proceedings of the 16th International Conference on Computational Linguistics*, Copenhagen, Denmark, August.
- M. Poesio, H. Cheng, R. Henschel, J. Hitzeman, R. Kibble, and R. Stevenson. 1999a. Specifying the parameters of centering theory: A corpus-based evaluation using text from application-oriented domains. In *Booktitle*, page Pages, Address, Month.
- M. Poesio, R. Henschel, J. Hitzeman, R. Kibble, S. Montague, and K. van Deemter. 1999b. Towards an annotation scheme for noun phrase generation. In *Booktitle*, page Pages, Address, Month.
- R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman Publishers.
- Ehud Reiter and Robert Dale. 1997. Building applied natural-language generation systems. *Journal of Natural-Language Engineering*, 3:57–87.
- Candace L. Sidner. 1983. Focusing in the comprehension of definite anaphora. In M. Brady and R. Berwick, editors, *Computational Models of Discourse*, pages 267–330. MIT Press, Cambridge, MA.
- Marilyn A. Walker. 1998. Centering, anaphora resolution, and discourse structure. In Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors, *Centering in Discourse*. Oxford University Press, Cambridge, MA.
- Bonnie Webber. 1979. *A Formal Approach to Discourse Anaphora*. Garland, NY.
- Maria Wolters and Donna K. Byron. 2000. Prosody and the resolution of pronominal anaphora. In *COLING–2000: Proceedings of the 18th International Conference on Computational Linguistics*, Saarbruecken, Germany.
- C. Yeh and C. Mellish. 1997. An empirical study on the generation of anaphora in Chinese. *Computational Linguistics*, 23(1):169–190.