A Comparison of Two Techniques for Resolving Demonstrative Anaphora

David McKoskey *

May 12, 2007

Abstract

This paper reviews two studies that use different techniques for finding the referent of a demonstrative expression. [Eckert and Strube, 1999] use rules and [Rocha, 2004] uses patterns of adjacent words (collocation patterns). After a brief overview of each study, the techniques described there are applied to a novel corpus and the results examined.

*This Plan B paper is submitted in partial fulfillment of the requirements for the degree of Masters of Arts in Linguistics at the University of Minnesota.

Contents

List of Figures

1 Introduction

I think the most interesting aspect of computational linguistics (CL) is that it is, by nature, an interdisciplinary field. Theories and methodologies are drawn from its two parent fields: linguistics and computer science. The CL texts I have surveyed ([**?**] [Jurafsky and Martin, 2000] [Manning and Schütze, 1999]) have all taken the same approach: CL research begins by stating a problem, and then setting out to create an algorithm or application to solve it. The problem statement is typically based on a representative sample of linguistic data or a corpus survey [1]. For instance, I will state the problem to be that of finding the referent of a demonstrative pronoun, such as the boxed forms in figure 1. Example 1 shows an instance where the referent of "that" is the activity of Dan buying Maria flowers. In example 2, the referent of "these" is a collection of entities: flowers of various type. In example 3, the demonstrative referent is a single entity: A. Johnson & Sons flower shop. Example 4 shows an instance where the referent is a single entity: "Dan and Maria's wedding anniversary" [2].

(1) Dan bought Maria a huge bouquet of flowers. That was really sweet.

(2) Dan had gone to three florists and looked at roses, daisies, and begonias. None of these were what he was looking for.

(3) Dan finally went to A. Johnson & Sons floral shop. That was where he found the right bouquet for Maria.

(4) Dan and Maria's wedding anniversary is coming up next week. This is her chance to surprise him.

Figure 1: Examples of Demonstrative Anaphora

Each example in figure 1 requires a different kind of linguistic knowledge to find the referent of the demonstrative pronoun. In example 1, one must recognize that the referent is an activity described in the previous sentence. Example 2 shows that a set of entities (flowers) must be identified in the preceding sentence, and the number and type of entities in that set must be decided. In example 3, the demonstrative pronoun's referent is the florist shop, which is the last entity mentioned in the preceeding sentence, while in example, 4, the referent is the occasion of Dan and Maria's Anniversary, which is the first entity mentioned in the preceding sentence. Taken together, how can these strategies be considered coherently? Is there any one procedure for resolving demonstrative anaphora that accounts for all of them? These are difficult questions to answer, but an attempt at answering them, however parsimoniously, must be attempted. For the computational linguist, one type of attempt is the development of a model.

Examples such as those in figure 1 guide the design for the algorithm or application that will solve the stated problem. However, within the design is another, stickier choice: the algorithm method that will be implemented. Multiple designs may be appropriate for the same problem. For instance, anaphora resolution systems have been built using heuristics [**?**], rules [Stuckardt, 2001], statistics [Ge et al., 1998], and machine learning [Soon et al., 2001]. Is one method better than another for this task? How does one know? How does one compare methods?

Taken in a general sense, questions as broad as these are beyond the scope of this study. What I will

[1] Please see the glossary in section A for the definitions of terms used throughout this paper.

[2] Special thanks to Professors Maria Gini and Dan Boley for providing the inspiration for this example.

do, however, is compare two studies by examining in detail their methods for resolving demonstrative anaphora. The first study is by [Eckert and Strube, 1999], who creates and tests a rule-based algorithm. The second is by [Rocha, 2004], who attempts to model demonstrative pronominal reference but does not present an algorithm. I will show how the results of Rocha's study could be used in an application that resolves demonstrative anaphora. Sections 2 and 3 will summarize the [Eckert and Strube, 1999] and [Rocha, 2004] studies and present their original results. In section 4, I test their techniques on a novel corpus (a collection of abstracts from the U.S. Department of Energy). A discussion is presented in section 5.

2 Eckert & Strube, 1999

At the 1999 EACL, Miriam Eckert and Michael Strube presented their work on a rule-based algorithm designed to identify the referent of a demonstrative pronoun [Eckert and Strube, 1999]. The algorithm developed by Eckert & Strube is based on Functional Centering [Strube and Hahn, 1999], a theory of attention based on Centering Theory [Grosz et al., 1995]. Before describing Eckert & Strube's algorithm, allow me to describe Centering Theory and Functional Centering.

Centering theory begins with the observation that "certain entities mentioned in an utterance are more central than others." Centering theory relates the choice of referring expression to the speaker and hearer's focus of attention and the perceived coherence of the discourse. One important assumption of Centering Theory is that discourse can be divided into segments. Coherence is considered differently whether global (between segments) or local (within segments). Centering Theory concerns only local coherence, though global coherence is considered to be different than the sum of all local coherences. Entities mentioned in a discourse segment, or centers, are ranked according to salience within a discourse segment. For a given discourse segment U_n, there is a set of candidate centers of attention called forward centers: $C_f(U_n)$. By definition, the most salient, or highest ranked center in $C_f(U_n)$ is the preferred center or $C_p(U_n)$. The center in U_n that is most salient in the *previous* segment U_{n-1} is the backward-looking center of U_n or $C_b(U_n)$. The relationship between $C_b(U_n)$ and $C_p(U_n)$ (they may or may not be the same entity) is an indication of perceived discourse coherence.

Ranking criteria for forward-looking centers varies, but one strong criterion cited is grammatical role: Subject > Direct Object > Other [Grosz et al., 1995]. Figure 2 shows an example of centering theory. Example 5 is more coherent than example 6 because example 5 centers on a single individual, and the backward-looking center and the preferred center refer to the same entity throughout. In 5, "John" is introduced as the subject of the sentence, and all pronouns in subject position in subsequent sentences refer to John. Example 6 is less coherent because the center of attention flips back and forth between two entities; John and the piano store. In 6.1, the subject and C_p is "John". In 6.2, the C_p shifts to the piano store, though the C_b is still "John". This pattern is reversed between sentences 6.2 and 6.3. Shifting centers between sentences continues throughout example 6. In fact, the backward-looking center and preferred center in example 6 never refer to the same entity.

Functional Centering [Strube and Hahn, 1999] is an extension of Centering Theory [Grosz et al., 1995]. Functional Centering begins with the observation that for free-word-order languages such as German, syntactic or grammaticality criteria are insufficient for ordering centers by salience. Instead, the "functional information structure" is used, which essentially tests whether an entity in a discourse segment is newly introduced ("discourse new") or has already been previously mentioned in a previous segment ("discourse old"). Related to the "discourse old / new" distinction is "cognitive load". Entities that require the least amount of inference to refer to are said to be lower in cognitive load than their inferred counterparts and are therefore higher ranking. Taken together, these two criteria form the core salience metric for Functional

(5) 1. John went to his favorite music store to buy a piano.
2. He had frequented the store for many years.
3. He was excited that he could finally buy a piano.
4. He arrived just as the store was closing for the day.

(6) 1. John went to his favorite music store to buy a piano.
2. It was a store he had frequented for many years.
3. He was excited that he could finally buy a piano.
4. It was closing just as John arrived.

Figure 2: Examples Illustrating Centering Theory from [Grosz et al., 1995]

Centering.

One additional feature of functional centering is that backward-looking centers are not used. Functional Centering "requires neither a backward-looking center, nor transitions, nor transition ranking criteria for an aphora resolution" [Strube and Hahn, 1999]. The reason is that information structure from the previous discourse segment ("discourse old") is already represented and has higher rank than new information.

Figure 3 shows an example of Functional Centering. In example 7, *building* is introduced as a brand-new (BN) or discourse-new entity. In example 8, the definite NP *the building* refers to the discourse entity from example 7, as does the pronoun *it* in example 9. Both are considered evoked (E), which is a lighter cognitive load than other referents.

(7) He lived his final nine years in a [rent-subsidized *building*] $_{BN}$ constructed especially for elderly survivors.

(8) When the [*building*]$_E$ opened, there was a waiting list.

(9) Once, [*it*]$_E$ held 333 survivors.

Figure 3: An Example of Cognitive Load from [Strube and Hahn, 1999]

Both Centering Theory and Functional Centering have provided solid theoretical footing for studies on pronoun resolution ([Byron and Stent, 1998] [Hahn and Strube, 1997] [Mitkov, 1998] inter alia). Pronoun resolution applications that implement these theories have tended to favor third-person personal pronouns ("he", "she", "it", "them", and "they"), which often refer to entities with the highest levels of salience [Gundel et al., 1993]. Software applications that implement Centering Theory or Functional Centering usually include at least the following components:

1. An operational definition of "discourse segment". Often, such a definition will leverage the structure

of the corpus used for the study. For instance, discourse segments in a dialogue may be changes in speaker "turns" [Sinclair, 1992] while in prose, it may be paragraphs or sentences [Webber, 1991].

2. A method for segmenting discourse, based on the definition above.

3. A functional representation of "discourse center". In centering theory, centers are semantic entities. In software systems, centers must be defined in a way that can be encoded in a computer's memory. Encoded centers may include (though not exhaustively) words, word sequences, noun phrases, syntax tree subtrees, clauses, or entities extracted from discourse.

4. A queue (ordered list) for storing representations of discourse centers. This queue is referred to by [Strube, 1998] and [Eckert and Strube, 1999] as an S-List.

5. A procedure for ordering the above queue, including ordering criteria.

2.1 Motivation

(10) Boeing ought to hire **him**$_i$ and give **him**$_i$ a junkyard$_j$ . . . . and see if **he**$_i$ could build a Seven Forty-Seven out of **it**$_j$. (sw2102)

(11) Now why didn't she [take him over there with her]$_i$? No, she didn't do **that**$_i$. (sw4877)

(12) I think the Soviet Union knows what we have and knows that we're pretty serious and if **they** ever tried to do anything, we would, we would be on the offensive. (sw3241)

(13)
B.29: I mean, the baby is like seventeen months and she just screams.
A.30: Uh-huh.
B.31: Well even if she knows that they're fixing to get ready to go over there.
They're not even there yet -
A.32: Uh-huh.
B.33: -you know.
A.34: Yeah. **It**'s hard.

(14) **It**'s hard to realize, that there are places that are just so, uh, bare on the shelves as there. (sw2403)

Figure 4: Examples of Anaphor Types from [Eckert and Strube, 1999]

Strictly speaking, the stated problem in [Eckert and Strube, 1999] is not demonstrative anaphora resolution, but finding the referent of a discourse deictic anaphor (an anaphor with a non-NP referent). Discourse deictic anaphors are often demonstrative pronouns, but not exclusively.
Figure 5 illustrates a few examples of discourse deixis. In each case, the referent of the boxed pronoun is a non-NP. In example 15, the referent of "It" is the event of John falling off his bicycle. The referent of "that" in example 16 is the fact that the cat disturbed a birthday cake. In example 17, the pronoun's referent is the invasion of Iran (an NP, but not listed as such in the discourse), a hypothetical situation or proposal.

The examples in figure 4 illustrates different types of anaphors that are marked in the TRAINS93 corpus and used to test Eckert & Strube's algorithm. Only discourse-deictic anaphors such as those illustrated

(15) John fell off his bike. It happened yesterday.

(16) The cat got into the birthday cake again. That's why I put him outside.

(17) The president wants to invade Iran. This is unacceptable course of action.

Figure 5: Examples of Discourse Deixis

in example 11 are analyzed. Other types include individual anaphors (example 10), which have NP antecedents, anaphors whose referents are inferred (example 12) or anaphors that refer to the general topic of conversation (example 13). Explitives (example 14) were also not considered, as they were considered "non-referring".

2.2 Dialogue Segmentation

Eckert & Strube chose the LDC TRAINS93 [**?**] to test their algorithm. Consistent with this choice, the authors also employ the definition of "discourse segment" for dialogues put forth by John Sinclair [Sinclair, 1992]: a discourse segment is a dialogue "synchronizing unit". A synchronizing unit is defined as single pair of dialogue initiation (I) and acknowledgement (A). In lay terms, a synchronizing unit is whenever a speaker takes a "turn" at speaking and then gives the other speaker an opportunity to do the same. An "initiation" requires an "acknowledgement" by the hearer, usually in the form of an affirmation ("right", "yes", "uh-huh", etc.).

When populating the algorithm's S-List, only those entities mentioned in an acknowledged initiation are added. Entities mentioned in unacknowledged initiations are ignored. Figure 6 (taken from [Eckert and Strube, 199 provides a general feel for how entities are selected for the S-List at different stages in the algorithm's execution. Using Strube & Hahn's recency criteria for ordering the S-List, entities are added to the back of the queue during an Initiation and discarded from the queue if unacknowledged. At the end of an Initiation, only the entity at the front of the queue requires acknowledgement. If unacknowledged, the first entity is removed leaving the second-highest entity as the new front of the queue.

Figure 6 provides an example of how the algorithm's S-List is populated. S1 is a conversation initiation, which adds one entity, "five boxcars of oranges" to the S-List. The acknowledgement SU2 is also an initiation, so no entities are removed. Instead, one is added: "Corning". The acknowledgement "um" in SU2 signifies a "turn" and pulls the item "5 boxcars of oranges" from the top of the stack. The entity "Corning" stays in the S-List because it was introduced with an acknowledgement. An additional initiation adds "orange warehouse" to the S-List. In initiation SU3, "five boxcars of oranges" is reintroduced, but at a lower rank due to its recency. The result is that the entity "Corning", having a higher rank and higher salience, is assigned to the pronoun "there" in SU3.

2.3 Eckert & Strube's Algorithm

In their algorithm, Eckert & Strube maintain the dialogue segmentation criteria outlined in section 2.2, figure 6, but change the ordering criteria and with it, the criteria by which items are removed from the

SU1	I	s:	so there - the five boxcars of oranges < sil > that are at -
			S-List: [5 boxcars of oranges]
SU2	A/I	u:	at < sil > - at Corning
			S-List: [5 boxcars of oranges, Corning]
	A	s:	um
-	I	u:	okay the orange warehouse < sil > um I have to
			S-List: [Corning, orange warehouse]
SU3	I	s:	you need - you need to get five < sil > five boxcars of oranges **there**
			S-List: [Corning, orange warehouse, 5 boxcars of oranges]

Figure 6: An Example of Functional Centering from [Strube and Hahn, 1999]

I-Incompatible	**A-Incompatible**
In an equating construction where a pronominal referent is equated with an abstract object	In an equating constructions where a pronominal referent is equated with a concrete individual object
In a copula construction whose adjectives are applied to abstract objects	In a copula construction whose adjectives are applied to individual objects
An argument of a verb describing an attitude that only takes an S' complement	An argument of a verb that describes physical contact / stimulation (non-metaphorically)
The object of *do*	
A "reason" for something	

Figure 7: Entity Classification Rules from [Eckert and Strube, 1999]

S-List. When a discourse-deictic pronoun is identified, the pronoun is evaluated using A-Incompatible or I-Incompatible criteria to determine the sort of referent it will likely have. I-Incompatible and A-Incompatible criteria are listed in figure 7. These criteria are based on the notion of compatibility between anaphor and antecedent:

> In the algorithm, we make use of the notion of anaphor-antecedent *compatibility* to distinguish between discourse deictic and individual reference... where *I-Incompatible* means *preferentially* associated with abstract objects and *A-Incompatibility* means *preferentially* associated with individual objects.

The context ranking criteria in figure 8 determines which candidate referents are added to the S-List and their ordering.

1. Candidate referents that comply with the criteria that fit the anaphor (I-Incompatible or A-Incompatible, depending)
2. Candidate referents within the same dialogue initiation I.
3. Candidate referents in the previous dialogue initiation I and is in its rightmost main clause (and subordinated clauses to its right).
4. Candidate referents in the previous dialogue initiation I and is in its rightmost complete sentence (and subordinated clauses to its right).

Figure 8: Context Ranking rules from [Eckert and Strube, 1999]

The algorithm proceeds by first identifying demonstrative pronouns[3] and then classifying them with the conditional statement in figure 9[4]. For demonstrative pronouns, the classification process includes identifying the pronoun's referent with the *resolveDD* function (figure 10)[5]. Starting with the first element, *resolveDD* attempts to co-index the demonstrative pronoun with an element in the context ranking (*co-index* is a function that co-indexes a pronoun with its referent). If the pronoun is an argument of *do*, then *resolveDD* co-indexes the pronoun with the VP of the element. Otherwise, *resolveDD* co-indexes the demonstrative pronoun with the element. As soon as the co-indexing succeeds, *resolveDD* exits the loop by returning "true". If co-indexing fails, "false" is returned.

Figure 11 illustrates how the algorithm is executed. In Example 19, the algorithm determines that the pronoun *it* is I-incompatible, as it is the argument of the verb *do*. No items in the previous sentence are I-Incompatible, nor are there any clauses to the left of the anaphor. The context ranking adds the last complete sentence in 18 to the S-List and co-indexes *it* to the VP-referent of the sentence in 18 ("if we went and policed").

2.4 Results of the Original Study

In Eckert & Strube's original study, five dialogues from the Switchboard corpus were selected and annotated to mark referents by anaphor type. Annotating pronoun types was done to test the accuracy of the

[3] Admittedly, identifying the use of "this" or "that" as pronominal is a complex process. I will assume that during the execution of this algorithm, demonstrative pronouns are correctly identified.

[4] A similar conditional statement is used to classify personal pronouns. Please see [Eckert and Strube, 1999] for details.

[5] The *resolveInd* function was not included in [Eckert and Strube, 1999].

1. **if** (DEM is I-incompatible)
   then if *resolveDD*(DEM)
      then classify as *DDDem*
      else classify as *VagDem*
2. **else if** (DEM is A-incompatible)
      then if *resolveInd*(DEM)
         then classify as *IDem*
         else classify as *VagDem*
3.   **else if** *resolveDD*(DEM)
         then classify as *DDDem*
4.   **else if** *resolvelnd*(DEM)
            then classify as *IDem*
            else classify as *VagDem*

Figure 9: Demonstrative Pronoun Detection Algorithm from [Eckert and Strube, 1999]

resolveDD(PRO) :=
1. **foreach**[a] element of context ranking **do**
2.   **if** (PRO is argument of *do*)
        then if (*co-index* PRO with VP of element)
          then add VP to S-List; **return** true
3.   **else if** (*co-index* PRO with element)
          then add element to S-List; **return** true
4. **return** false.

Figure 10: Function *resolveDD* from [Eckert and Strube, 1999]

[a]The "foreach" looping construct is used in many computer programming languages to iterate sequentially through a list of items. In this case, the looping ends early if the *co-index* function succeeds.

(18) I mean, if we went and policed, just like you say, every country when they had squabbles,

(19) Well, but we've done **it** before.

Figure 11: Algorithm Execution Example from [Eckert and Strube, 1999]

algorithm after execution. Two annotators were employed, who showed very good agreement (90%). Four anaphor types were identified and marked:

1. "Individual anaphors": pronouns with NP antecedents. Most often, third-person personal pronouns.
2. "Discourse deictic": pronouns that co-specify with non-NP constituents (VPs, sentences, strings of sentences. Most often, demonstrative pronouns.
3. "Inferrable Evoked pronouns": a plural pronoun that indirectly co-specifies with a singular antecedent.
4. "Vague": do not have a specific clause antecedent, referring instead to "the general topic" and "do not have a specific clause antecedent".

The algorithm was run "by hand", that is to say, it was not implemented as software. Not surprisingly, results were checked "by hand" as well. For many computational linguists (including Eckert & Strube), algorithm performance is measured two ways: precision and recall. "Precision" is the number of correctly identified referents across all discourse deictic expressions. "Recall" is the number of correctly identified referents where resolution is attempted. The difference between precision and recall is that some demonstrative expressions were found but their resolution not attempted. Precision for this study was reported at 65% and recall at 70%. Unfortunately, the original study did not report accuracy by anaphor type. Results were reported for each dialogue, though accuracy did not appear to differ a great deal between dialogues. The greatest number of errors was reported for candidate referents annotated as "vague".

3 Rocha, 2004

In his 1997 Ph.D. thesis, Marco Rocha outlines a multiple-strategy approach for finding the referent of demonstrative pronouns [Rocha, 1997]. In the study presented here ([Rocha, 2004]), the author summarizes this approach:

> A processing strategy is defined in the study as a description of a resolution path leading to the identification of the antecedent. Each Processing strategy has been given a name which sums up the essential aspects of the resolution process based on the analysis of corpus data.

There are five component strategies in Rocha's approach. The first two strategies, called "naive" approaches, are "First Candidate" and "First Candidate Chain". These techniques employ the approach developed by [Hobbs, 1978] and use syntactic information to locate a demonstrative pronoun's referent. According to the author, the success rate for this technique for demonstrative pronouns is fairly low: 28%. The remaining three strategies are "Discourse Knowledge", "Deixis", and "Collocation Knowledge". This study develops the "Collocation Knowledge" strategy.

Rocha defines "Collocation Knowledge" as "...a regular association between co-occurrence patterns of the anaphor and a resolution path for the identification of the antecedent". Rephrased slightly, a particular string of words is predictive of a particular heuristic or pattern that finds a demonstrative pronoun's referent. In figure (12), the collocation in 21 refers to the previous move. A "move" is the most recent speech act in the dialogue sequence made by the other participant [Sinclair, 1992]. The result is the resolution strategy "Select the previous move as the referent of the string 'something like that"'. This strategy is used for the remaining examples Rocha cites. For instance, the collocation patterns "this means..." and 'that's why..." are also shown to refer to the previous "move". In example 21, "that" refers to "he must have the idea that they wouldn't support it".

(20) Erm, they do support the humanities - but if Joe Eccleston hasn't mentioned them *he must have the idea that they wouldn't support it* or something.

(21) Something like **that**

Figure 12: Dialogue sample from [Rocha, 2004]

3.1 Findings in [Rocha, 2004]

The result of [Rocha, 2004] is a set of word collocations and their resolution paths, summarized in figure 13. In this table, the left column shows the collocation or word pattern, while the right column describes the steps necessary to find the collocation's referent. Results are shown for both English and Portuguese. For this particular study, Rocha reports the same resolution path for all collocations: "select the previous move as the antecedent of the demonstrative pronoun".

Collocation	**Antecedent**
"something / anything / nothing like that"	Select the previous move as the antecedent.
"this / that's why"	Select the previous move as the antecedent.
"this means..."	Select the previous move as the antecedent.
"this / that X-be [a] (not) (ADV) right / wrong"	Select the previous move as the antecedent.
"por isso que X" meaning "for that which " (X is any complement following the conjunction)	Select the previous move as the antecedent.
"pelo menos isso"	Select the previous move as the antecedent. (Caveat: this path is based only on a single example)

Figure 13: Findings in [Rocha, 2004]

[a] X-be is any form of "be", and negation and an additional adverb are optional

Next, Rocha tests the strategies in figure 13 on some data left out of the corpus study that gave rise to them. The results are mixed; some collocation / resolution path pairs work well, some don't. In 60 analyzable cases of "something like that", 20 of the antecedents were the previous move, but 40 were a previous NP. He also reports problems delimiting the scope of NP antecedents, especially when they contained conjunctions ("and", "or", etc.). Rocha concludes that "It seems clear... that the collocation needs further investigation, since the proposed resolution path does not hold in any regular fashion." By contrast, the collocation "this is / that's why" showed that "the resolution path seems to work fairly well in the majority of cases analyzed, but it would require a little more sophistication in some situations."

3.2 Machine Learning Implementation

Noteworthy at this point is that unlike [Eckert and Strube, 1999], Rocha does not develop an algorithm for resolving demonstrative pronouns. The technique for finding collocations and their resolution paths, however, could be used to provide input for statistical modeling. Associations between patterns, such as those illustrated in figure 13 are good candidates for machine learning algorithms (algorithms that benefit from experience [Mitchell, 1997]). Supervised machine learning algorithms require an initial experience or "training period" during which an association (often stated statistically) is established between a particular set of initial conditions and a desired output.

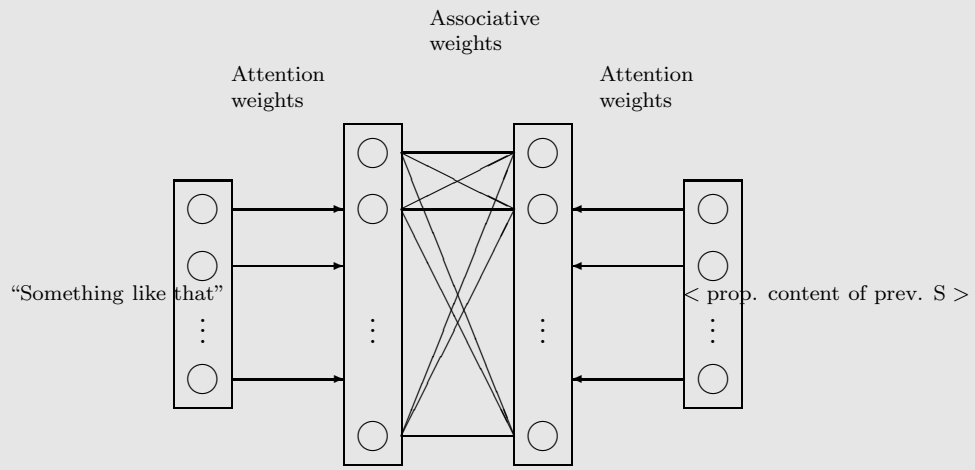

Figure 14: A Sample Neural Network Suitable for [Rocha, 2004]

A neural network would be a natural way to implement the results of Rocha's study algorithmically. In figure 12, an association is made between example 20 and example 21. This example illustrates at least two ways to train a neural network. One form of training is between string literals: the string "Something like that" is literally associated with the text in example 20 ("Erm,..."). Obviously, this approach is not very flexible and would likely require an enormous training set to yield even modest results. Far better would be to associate target strings with more general patterns that retain the predictive power of the association. For instance, an association between the string literal in example 21 ("Something like that") with the propositional content of the previous utterance would be a more flexible association, but would keep its predictive power.

Figure 14 illustrates how one could implement the results of Rocha's study as a neural network. On either side of the banks of internal nodes are the items to be associated. On the left is the string from example 21 and on the right is the resolution path, in the form of "< prop. content of prev. S >" (find the propositional content of the previous sentence). Trained with this pairing, the network learns to report that the referent of "that" is the propositional content of the previous sentence.

The example in 12 is used throughout his study, hence my use of it here. For use in a statistical model, however, the resolution path for "something like that" as simply "the previous move" would need further

refinement before training a statistical model.

4 Experiments Implementing [Eckert and Strube, 1999] and [Rocha, 2004]

To compare Eckert & Strube and Rocha's methods for resolving demonstrative anaphora, their techniques (outlined in sections 2 and 3) were implemented as algorithms and executed on novel data. The original test data used for Eckert & Strube and Rocha's studies was the TRAINS93 dialogue corpus. To ensure that these methods were not biased toward dialogues, a collection of Department of Energy (DOE) abstracts was used as test data. To encode Eckert & Strube and Rocha's techniques as algorithms, the application template shown in figure 15 was used. The "apply selection criteria" step is where either Eckert & Strube or Rocha's techniques were plugged-in. The observations below are based on a sample of 20 DOE abstracts that were run "by hand" through the template application (figure 15).

1. **Input text**: text files containing DOE abstract data are opened and processed a line at a time.
2. **Segment text**: input text is separated into documents, paragraphs, sentences, and words.
3. **Search for demonstrative pronoun**: as each line of text is input, the line is searched for a demonstrative pronoun[a]. If a demonstrative pronoun is found, go to step 4. Otherwise, return to step 1.
4. **Determine context size**: determine the size of the context for extracting candiate referents. This study assumes that only referents *in the text* can be candidates.
5. **Extract potential referents**: parse all discourse segments syntactically and copy segments from them to represent candidate referents. Most often, candidates are NPs and full sentences.
6. **Apply selection criteria**: apply criteria for selecting the referent of a demonstrative expression, shown in figure 16 and 17.
7. **Repeat**: if more text remains, return to step 1.

Figure 15: Algorithm Template to Implement [Eckert and Strube, 1999] and [Rocha, 2004]

[a]Admittedly, determining the use of a demonstrative form is not a simple task. If the word immediately after a demonstrative form is an NP, the form is used as a determiner. The form "that" can also be used as a complementizer and relativizer; determining whether or not "that" is a pronoun often requires a full parse of the sentence

4.1 Test Data

Figures 18 and 19 show the full set of data used to test the algorithm implementations outlined in sections 2 and 3 (above). In each example, the demonstrative pronoun is bounded by a box and its referent is italicized. Sections 4.3 and 4.4 will use this data.

4.2 Test Implementation for [Eckert and Strube, 1999]

In Eckert & Strube's original study, the "context ranking" criteria span only the current and previous dialogue initiation (I). To simulate this using DOE abstracts, candidate referents added to the S-List were

1. Since context ranking rules concern only the current and previous sentence, the context for extracting candidate referents is only the current and previous sentence.
2. Determine the pronoun's compatibility class (A-Incompatible or I-Incompatible)
3. Using the context ranking criteria, add all potential referents to the S-List
4. Filter the S-List according to A-Incompatible or I-Incompatible criteria (Table 7).
5. Return the first element in the S-List

Figure 16: Referent Selection Criteria, per [Eckert and Strube, 1999]

1. The context for extracting candidate referents is essentially unbounded.
2. Filter the set with all known input patterns.
3. Return the referent with the strongest associative strength.

Figure 17: Referent Selection Procedure, per [Rocha, 2004]

drawn only from the current sentence (where the demonstrative form is found) and the previous. The adapted context ranking is illustrated in figure 20.

Context ranking rules (3) and (4) (figure 20) required some additional constraint. The definition of "rightmost main clause" can be interpreted in a way that leads to deadlocks. Figure 21 illustrates an example of deadlock when the scope of a candidate referent in the same sentence is ambiguous. The referent may not be the entire NP, but if not, how does one determine which part is? We cannot allow parts of an NP to be candidates because the algorithm does not provide a way of breaking deadlocks. As part of the experiment, candidate referents will be maximized for scope.

4.3 Results for the Test Implementation of [Eckert and Strube, 1999]

An application using the application template in figure 15 and the referent selection criteria in figure 16 (Eckert & Strube's algorithm) was tested on the data in figures 18 and 19. Diagrams detailing the execution of Eckert & Strube's algorithm on each example are available in Appendix B. For the sake of brevity, illustrative examples will be presented here. In all examples, the demonstrative pronoun meets the I-Incompatible criteria. All candidate referents were vetted using I-Incompatible criteria.

In each figure below, the results of executing Eckert & Strube's algorithm on a sample is presented as a table. Above the table is the sample text. The center column of the table contains S-List candidates. The rightmost column indicates the reason the candidate referent was chosen for the S-List, whether by I-Incompatible criteria or context ranking. The leftmost column indicates the outcome of the context ranking's decision: a check mark indicates that the correct referent was selected. An asterisk indicates that an incorrect choice or no choice was selected.

(22) *Eventually rat lenses irradiated with 1200 rad developed severe (4+) cataracts 32 weeks after exposure.* The authors claim this is the first report of cortical fibers increasing transiently in volume prior to cataract formation.

(23) Bifurcation behavior is examined, and dynamical behavior is numerically computed and compared to the experimental data. *A theory* is developed regarding the nature of the experimentally observed relaxation oscillations and spiking behavior based on the interaction of "weak" and "strong" modes, and this is demonstrated in the numerical simulations for two modes.

(24) Probability of relapse was 37% for patients in the MTX group and 70% for patients in the CSP group (p¡0.05). *Transplant-related deaths were more frequent in the MTX group and leukemic deaths more frequent in the CSP group* although this may have been related to an uneven distribution of high-risk patients. Long term disease-free survival was comparable.

(25) *The effectiveness of activated carbon for removal of PCBs from water is adversely affected by background humic substances.* Mechanisms responsible for this may be competition and/or complexation between humic substances and PCB molecules.

(26) It was shown that *preparations of bovine kidney gamma-glutamyl transferase of differing degrees of purity are phosphorylated by cAMP-dependent protein kinase.* This is accompanied by a decrease in both the transferase and hydrolase activities of the enzyme.

(27) This phenomenon may be due to interaction between the free PCB molecules and more strongly adsorbed calcium-humate-PCB complexes. *Removal of trace cadmium, lead, and zinc* by carbon-immobilized 8-hydroxyquinoline is far superior to that of virgin carbon.

(28) This paper describes the results of field testing of a process for removing gases from geothermal steam, based on condensation and re-evaporation of the steam in a shell-and-tube heat exchanger, upstream from the generator. The gases that are removed are blown off with a small amount of the steam fed in the process, and the final product is *steam with a low concentration of Hsub 2S and COsub 2*, at a pressure slightly lower than that of the initial steam.

(29) It was found that PMPSPn becomes cross-linked by the action of radiation. It was shown that the *radio-resistance* of this polymer is increased in comparison with that of polyorganosiloxanes with no silaoxadihydrophenanthrenyl units in the side chain.

(30) 14 MeV have been measured using the Livermore and Ohio University neutron time-of-flight facilities. *The data* are reviewed and compared with that of other workers.

(31) The morphological alterations were similar for both types of insult, but the time sequences of events were quite different. *The post-irradiation recovery period* was vastly extended compared with that of the hydroxyurea treated hair roots.

Figure 18: DOE Abstract Samples Used for Testing, Part 1

(32) *The factors that have changed public attitudes towards the use of nuclear power* are reviewed. These are seen as the link between nuclear weapons and nuclear power, radiation leaks and reactor accidents and the uneconomic nature of nuclear power.

(33) *A number of the statistical procedures employed*, which have been prepared as independent computer codes, are also described. These are of general applicability in many areas of probabilistic assessment.

(34) *Six mistakes by the operators* led to the accident at the Cherobyl nuclear reactor. These have been studied.

(35) The steps taken to ensure that similar accidents could not occur in the UK are mentioned. However, in addition to efforts to prevent accidents, *emergency arrangements* have been made and these are explained and the chain of command for nuclear installations emergency arrangements given. (U.K.).

(36) At 150-180sup 0C *monomeric compounds* are formed; these are converted to polymeric products above 200sup 0C.

(37) Following the accident at Chernobyl, the nuclear policies of many *governments* have been reconsidered and restated. Those in favour of nuclear power are those with highly centralised state bureaucracies, such as France and the USSR, where public opinion is disregarded.

(38) The present paper reviews the results obtained for the influence of channeling and over-barrier motion on the relativistic electron and positron radiation in the crystal at different frequencies. *The predictions* of the Born theory of coherent radiation in the crystal are compared with those of the theory of radiation in channeling and over-barrier motion.

(39) We show from the comparison of the experimental angular distributions of one-neutron transfer reactions with that of the elastic scattering that a specific phase rule emerges. This elastic-transfer phase rule states that *the oscillations in the neutron transfer cross-section* should be in phase with those of the elastic cross-section.

(40) In the same period, about 250 recessive alleles would be eliminated in heterozygotes given 2.5% heterozygous disadvantage. These *deleterious heterozygous effects* should not be combined with those of dominants, as has been done in some previous risk estimates.

(41) Actuarial relapse and survival rates at 2 yr were 56% and 9.5% respectively. These *data* were not remarkably different from those in previous studies using 10 Gy of TBI administered as a single dose.

Figure 19: DOE Abstract Samples Used for Testing, Part 2

1. Candidate referents that comply with the criteria that fit the anaphor (I-Incompatible or A-Incompatible, depending)
2. Candidate referents within the same S.
3. Candidate referents in the previous S and is in its rightmost main clause (and subordinated clauses to its right).
4. Candidate referents in the previous S and is in its rightmost complete sentence (and subordinated clauses to its right).

Figure 20: Context Ranking Adapted from [Eckert and Strube, 1999]

This phenomenon may be due to interaction between the free PCB molecules and more strongly adsorbed calcium-humate-PCB complexes. *Removal of trace cadmium, lead, and zinc by carbon-immobilized 8-hydroxyquinoline* is far superior to that of virgin carbon.

	S-List	**Criterion**
?	Removal	Candidate referent, same S
?	Removal of trace cadmium, lead, and zinc	Candidate referent, same S
?	Removal of trace cadmium, lead, and zinc by carbon-immobilized 8-hydroxyquinoline	Candidate referent, same S

Figure 21: Example of a Deadlock Due to Classification Criteria

Figure 22 shows an example of a demonstrative pronoun whose referent is correctly identified as the previous sentence. The I-Incompatible criteria yield no candidates from the previous or current sentence. Entities mentioned previously in the same sentence fail to agree in number. There are no rightmost main clauses in the previous sentence because the previous sentence has only one clause. The result is that the previous sentence is the only candidate added to the S-List. In examples 27, 30, 31 and 37, the referent is correctly identified, while in examples 38 and 39 it is not. This strategy fails when the referent is only one entity mentioned in the previous sentence, rather than the propositional content of the previous sentence.

Eventually rat lenses irradiated with 1200 rad developed severe (4+) cataracts 32 weeks after exposure. The authors claim this is the first report of cortical fibers increasing transiently in volume prior to cataract formation.

Result	S-List	Criterion
✓	Eventually rat lenses irradiated with 1200 rad developed severe (4+) cataracts 32 weeks after exposure	Rightmost complete sentence, previous S

Figure 22: Referent Correctly Identified as Previous Sentence

Figure 23 (same as example 42) shows a case where the only candidate referent is the previous sentence. This would be an erroneous result, since the correct referent is one entity mentioned in the previous sentence, not the entire sentence. However, this candidate is rejected prior to consideration because the event or proposition of an entire sentence is considered *singular* in number, whereas the demonstrative pronoun is plural. This instance of algorithm execution fails because there are no candidates to choose from when the correct referent is present within the discourse segment.

Following the accident at Chernobyl, the nuclear policies of many *governments* have been reconsidered and restated. Those in favour of nuclear power are those with highly centralised state bureaucracies, such as France and the USSR, where public opinion is disregarded.

Result	S-List	Criterion
*	(none)	NA

Figure 23: An S-List With No Candidates

Figure 24 shows a demonstrative pronoun whose referent is in the same sentence. Again, the I-Incompatible criteria fail to add any candidate referents from the previous sentence, which also has only one clause. The candidate referent is correctly chosen because it is mentioned early in the sentence. Figures 28, 29, 32, 34, 35, 36, 43, 45 and 46 all show these results, while figures 33, 40 and 44 use the same strategy but are not successful. This strategy fails when the referent is not the first candidate in the same sentence as the demonstrative pronoun.

Bifurcation behavior is examined, and dynamical behavior is numerically computed and compared to the experimental data. *A theory* is developed regarding the nature of the experimentally observed relaxation oscillations and spiking behavior based on the interaction of "weak" and "strong" modes, and this is demonstrated in the numerical simulations for two modes.

Result	S-List	Criterion
✓	A theory	Candidate referent in same S
	The nature of experimentally observed relaxation oscillations and spiking behavior based on the interaction of "weak" and "strong" modes	Candidate referent in same S

Figure 24: Referent Correctly Identified as an Entity in the Previous Sentence

4.4 Implementation and Results for [Rocha, 2004]

Implemented properly, pairs of demonstrative pronouns and resolution paths would be extracted from a large corpus of data, not twenty examples. Twenty examples is too few. However, if we assume like Rocha that the resolution path is more than string matching, the pairs we find in examples 22 through 41 may still lend us some insight into how Rocha's technique would perform if scaled up.

First, the demonstrative pronouns and their resolution paths were identified for all examples in figures 18 and 19 These pairs are shown in figure 25. Next, an application using the template in figure 15 and the referent selection criteria in figure 17 (the pairs extracted previously) was tested on the data in figures 18 and 19.

5 Discussion

I found that Eckert & Strube's algorithm for finding the referent of a discourse-deictic pronoun worked for most of the sample data (section 4.1). The authors' original results (precision = 65%, recall = 70%) are by no means stellar results, but are similar to the results I obtained in section 4.3: 75%. The chief problem that stood in the way of better results was that the I-Incompatible criteria as stated by [Eckert and Strube, 1999] were never useful for finding candidate referents. In all examples (figures 18 and 19), they simply did not apply. This result suggests that Eckert & Strube's I-Incompatible criteria were formulated for the data sample they used for testing (TRAINS93). If one were to properly reformulate the I-Incompatible criteria to find candidate NP referents in DOE abstracts, resolution accuracy would probably increase.

Future work on Eckert & Strube's algorithm will also require an explicitly stated solution to the problem of deadlocks (figure 21). The task of inferring the size or scope of an entity based on its text description is a common problem, and anyone who reads text regularly does so with ease. Example 2 requires the reader to group the three types of flowers Dan surveyed in order to understand that none met with his approval. For anyone with a fair grasp of English, doing so is not difficult. Encoding this process as an algorithm, however, does not strike me as a simple task.

Ex.	Pronoun	Resolution Path
22	this is	The propositional content of the previous sentence
23	this is	The subject of the same sentence
24	this may	The propositional content of the previous sentence
25	this may	The propositional content of the previous sentence
26	this is	The rightmost main clause of the previous sentence
27	that of	The subject of the same sentence
28	that of	A non-parallel NP that is also the last NP prior to the pronoun
29	that of	A non-parallel NP that is two NP's prior to the pronoun
30	that of	The subject of the same sentence
31	that of	A non-parallel NP that is the subject of the same sentence
32	these are	The subject of the previous sentence
33	these are	The subject of the previous sentence
34	these have	The subject of the previous sentence
35	these are	A non-parallel NP that is the last NP prior to the pronoun
36	these are	The subject of the previous sentence
37	those with	A non-parallel NP that is the last NP prior to the pronoun
38	those of	A non-parallel NP that is the subject of the same sentence
39	those of	A non-parallel NP that is the last NP prior to the pronoun
40	those of	A non-parallel NP that is the subject of the same sentence
41	those in	A non-parallel NP that is the subject of the same sentence

Figure 25: Demonstrative Pronouns and their Resolution Paths

The results of my tests using Rocha's techniques are similar to those obtained by Rocha in his original study (section 3.1). No single resolution path garnered from the examples in figures 18 and 19 works in all situations or for all pronouns. Further testing with a (much) larger language sample should be done. The number of examples used in this study is simply too small to make statistical predictions. For instance, Examples 37 through 41 concern the pronoun "those". It would be unsafe to predict a 60% accuracy rate for the resolution path "A non-parallel NP that is the subject of the same sentence".

I also found that Rocha's observation that some resolution paths "require a little more sophistication in some situations" is apt for my results as well. Examples 39 and 40, among others, have the same collocation sequence (pronoun + following word) but have different resolution paths. Some other discriminant must be added to these resolution paths before they can be used in a statistical model. As was true for Rocha's findings, these resolution paths need further refinement. For instance, longer collocations and more precise characterization of referent attributes would make a statistical model more accurate for easy cases.

A Glossary

Term	Definition
Algorithm	A series of steps, typically using iteration or recursion, to systematically define, deconstruct, and solve a problem. Algorithms require or assume certain pre-conditions and return pre-determined results.
Corpus	A corpus is a collection of data, typically in a pre-determined form. The Switchboard corpus, for instance, id a collection of conversations, annotated to indicate which of two speakers is speaking. The Penn Treebank is a collection of airline requests annotated with syntactic constituents. Corpora are often used as training and test data for applications or algorithms.
Discourse Deictic	Having a non-NP constituent or referent.
Discourse Segment	A defined chunk of discourse the author of the study is considering. A discourse segment may be a sentence, a paragraph, a "turn" in a dialogue [Sinclair, 1992], &c. Any author using the term must carefully define it.
Machine Learning	a class of algorithms that improve with experience. Machine learning algorithms are divided into two classes: supervised and unsupervised. Supervised algorithms require training data, while unsupervised ones do not.
NP	Abbreviation for "noun phrase".
VP	Abbreviation for "verb phrase".
Training	Supervised machine learning algorithms require a pairing of input data and "correct" output data to train on. Often, the result of training is a statistical model of the data.
Testing	Supervised algorithms also require that roughly %25 of the available training data to be set aside for testing the algorithm. Often more than one round of testing is performed, in which %75 of the data is chosen at random for training, and the remainder used for testing.

Figure 26: Glossary

B [Eckert and Strube, 1999] Algorithm Execution Results (Section 4.3)

This section includes details on the execution of the [Eckert and Strube, 1999] on the test examples in figures 18 and 19. In each example, the demonstrative pronoun is bounded by a box and its correct referent is italicized. Each table contains the contents of the S-List, with the proposed referent at the top of the list. A check mark next to a candidate referent indicates success. An asterisk indicates that the wrong referent was chosen.

Eventually rat lenses irradiated with 1200 rad developed severe (4+) cataracts 32 weeks after exposure. The authors claim this is the first report of cortical fibers increasing transiently in volume prior to cataract formation.

Result	S-List	Criterion
✓	Eventually rat lenses irradiated with 1200 rad developed severe (4+) cataracts 32 weeks after exposure	Rightmost complete sentence, previous S

Figure 27: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 22

Bifurcation behavior is examined, and dynamical behavior is numerically computed and compared to the experimental data. *A theory* is developed regarding the nature of the experimentally observed relaxation oscillations and spiking behavior based on the interaction of "weak" and "strong" modes, and this is demonstrated in the numerical simulations for two modes.

Result	S-List	Criterion
✓	A theory	Candidate referent in same S
	The nature of experimentally observed relaxation oscillations and spiking behavior based on the interaction of "weak" and "strong" modes	Candidate referent in same S
	Bifurcation behavior is examined, and dynamical behavior is numerically computed and compared to the experimental data.	Rightmost complete sentence, previous S

Figure 28: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 23

Probability of relapse was 37% for patients in the MTX group and 70% for patients in the CSP group (pj0.05). *Transplant-related deaths were more frequent in the MTX group and leukemic deaths more frequent in the CSP group* although this may have been related to an uneven distribution of high-risk patients. Long term disease-free survival was comparable.

Result	S-List	Criterion
✓	Transplant-related deaths were more frequent in the MTX group and leukemic deaths more frequent in the CSP group	Candidate referent in same S

Figure 29: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 24

The effectiveness of activated carbon for removal of PCBs from water is adversely affected by background humic substances. Mechanisms responsible for this may be competition and/or complexation between humic substances and PCB molecules.

Result	S-List	Criterion
✓	The effectiveness of activated carbon for removal of PCBs from water is adversely affected by background humic substances	Rightmost complete sentence, previous S

Figure 30: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 25

It was shown that *preparations of bovine kidney gamma-glutamyl transferase of differing degrees of purity are phosphorylated by cAMP-dependent protein kinase.* This is accompanied by a decrease in both the transferase and hydrolase activities of the enzyme.

Result	S-List	Criterion
✓	preparations of bovine kidney gamma-glutamyl transferase of differing degrees of purity are phosphorylated by cAMP-dependent protein kinase	Rightmost complete sentence, previous S

Figure 31: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 26

This phenomenon may be due to interaction between the free PCB molecules and more strongly adsorbed calcium-humate-PCB complexes. *Removal of trace cadmium, lead, and zinc by carbon-immobilized 8-hydroxyquinoline* is far superior to that of virgin carbon.

Result	S-List	Criterion
✓	Removal of trace cadmium, lead, and zinc by carbon-immobilized 8-hydroxyquinoline	Candidate referent in same S
	This phenomenon may be due to interaction between the free PCB molecules and more strongly adsorbed calcium-humate-PCB complexes	Rightmost complete sentece, previous S

Figure 32: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 27

This paper describes the results of field testing of a process for removing gases from geothermal steam, based on condensation and re-evaporation of the steam in a shell-and-tube heat exchanger, upstream from the generator. The gases that are removed are blown off with a small amount of the steam fed in the process, and the final product is steam with a low concentration of Hsub 2S and COsub 2, at a *pressure* slightly lower than that of the initial steam.

Result	S-List	Criterion
*	The gases that are removed	Candidate referent in same S
	a small amount of the steam fed in the process	Candidate referent in same S
	the final product	Candidate referent in same S
	steam with a low concentration of Hsub 2S and COsub 2	Candidate referent in same S
	(steam) pressure	

Figure 33: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 28

It was found that PMPSPn becomes cross-linked by the action of radiation. It was shown that the *radio-resistance of this polymer* is increased in comparison with that of polyorganosiloxanes with no silaoxadihydrophenanthrenyl units in the side chain.

Result	S-List	Criterion
✓	radio-resistance of this polymer	Candidate referent in same S
	It was found that PMPSPn becomes cross-linked by the action of radiation.	Rightmost complete sentence, previous S

Figure 34: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 29

14 MeV have been measured using the Livermore and Ohio University neutron time-of-flight facilities. *The data* are reviewed and compared with that of other workers.

Result	S-List	Criterion
✓	The data	Candidate referent in same S
	14 MeV have been measured using the Livermore and Ohio University neutron time-of-flight facilities	Rightmost sentence, previous S

Figure 35: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 30

The morphological alterations were similar for both types of insult, but the time sequences of events were quite different. *The post-irradiation recovery period* was vastly extended compared with that of the hydroxyurea treated hair roots.

Result	S-List	Criterion
✓	The post-irradiation recovery period	Candidate referent, same S
	The morphological alterations were similar for both types of insult, but the time sequences of events were quite different.	Rightmost complete sentence, previous S

Figure 36: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 31

The factors that have changed public attitudes towards the use of nuclear power are reviewed. These are seen as the link between nuclear weapons and nuclear power, radiation leaks and reactor accidents and the uneconomic nature of nuclear power.

Result	S-List	Criterion
✓	The factors that have changed public attitudes towards the use of nuclear power	Rightmost complete sentence, previous S

Figure 37: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 32

A number of the statistical procedures employed, which have been prepared as independent computer codes, are also described. These are of general applicability in many areas of probabilistic assessment.

Result	S-List	Criterion
*	A number of the statistical procedures employed, which have been prepared as independent computer codes, are also described.	Rightmost complete sentence, previous S

Figure 38: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 33

Six mistakes by the operators led to the accident at the Cherobyl nuclear reactor. These have been studied.

Result	S-List	Criterion
*	Six mistakes by the operators led to the accident at the Cherobyl nuclear reactor.	Rightmost complete sentence, previous S

Figure 39: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 34

The steps taken to ensure that similar accidents could not occur in the UK are mentioned. However, in addition to efforts to prevent accidents, *emergency arrangements* have been made and these are explained and the chain of command for nuclear installations emergency arrangements given. (U.K.).

Result	S-List	Criterion
*	efforts to prevent accidents	Candidate referent in same S
	emergency arrangements	Candidate referent in same S

Figure 40: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 35

At 150-180sup 0C *monomeric compounds* are formed; these are converted to polymeric products above 200sup 0C.

Result	S-List	Criterion
✓	monomeric compounds	Candidate referent, same S

Figure 41: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 36

Following the accident at Chernobyl, the nuclear policies of many *governments* have been reconsidered and restated. Those in favour of nuclear power are those with highly centralised state bureaucracies, such as France and the USSR, where public opinion is disregarded.

Result	S-List	Criterion
*	(none)	NA

Figure 42: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 37

The present paper reviews the results obtained for the influence of channeling and over-barrier motion on the relativistic electron and positron radiation in the crystal at different frequencies. *The predictions* of the Born theory of coherent radiation in the crystal are compared with those of the theory of radiation in channeling and over-barrier motion.

Result	S-List	Criterion
✓	The predictions	Candidate referent, same S
	the Born theory of coherent radiation in the crystal	Candidate referent, same S

Figure 43: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 38

We show from the comparison of the experimental angular distributions of one-neutron transfer reactions with that of the elastic scattering that a specific phase rule emerges. This elastic-transfer phase rule states that *the oscillations in the neutron transfer cross-section* should be in phase with those of the elastic cross-section.

Result	S-List	Criterion
*	This elastic-transfer phase rule	Candidate referent, same S
	the oscillations in the neutron transfer cross-section	Candidate referent, same S

Figure 44: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 39

In the same period, about 250 recessive alleles would be eliminated in heterozygotes given 2.5% heterozygous disadvantage. *These deleterious heterozygous effects* should not be combined with those of dominants, as has been done in some previous risk estimates.

Result	S-List	Criterion
✓	These deleterious heterozygous effects	Candidate referent, same S

Figure 45: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 40

Actuarial relapse and survival rates at 2 yr were 56% and 9.5% respectively. *These data* were not remarkably different from those in previous studies using 10 Gy of TBI administered as a single dose.

Result	S-List	Criterion
✓	These data	Candidate referent, same S

Figure 46: Execution Details for [Eckert and Strube, 1999] Algorithm on Example 41

References

[Byron and Stent, 1998] Byron, D. and Stent, A. (1998). A preliminary model of centering in dialog. In *Proceedings of the 36 Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1475 – 1477.

[Eckert and Strube, 1999] Eckert, M. and Strube, M. (1999). Resolving discourse deictic anaphora in dialogues. In *Proceedings of the 9th Conference of the European Chapter of the Association for Computational Linguistics*, pages 37–44.

[Ge et al., 1998] Ge, N., Hale, J., and Charniak, E. (1998). A statistical approach to anaphora resolution. In *Proceedings of the Sixth Workshop on Very Large Corpora*.

[Grosz et al., 1995] Grosz, B., Joshi, A., and Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.

[Gundel et al., 1993] Gundel, J., Hedberg, N., and Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):274–307.

[Hahn and Strube, 1997] Hahn, U. and Strube, M. (1997). Centering in-the-large: computing referential discourse segments. In *Proceedings of the eighth proceedings on European chapter of the Association for Computational Linguistics*, pages 104–111.

[Hobbs, 1978] Hobbs, J. (1978). Resolving pronoun references. *Lingua*, 44:311–338.

[Jurafsky and Martin, 2000] Jurafsky, D. and Martin, J. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall.

[Manning and Schütze, 1999] Manning, C. and Schütze, H. (1999). *Statistical Approaches to Natural Language Processing*. Prentice-Hall.

[Mitchell, 1997] Mitchell, T. (1997). *Machine Learning*. McGraw Hill.

[Mitkov, 1998] Mitkov, R. (1998). Robust pronoun resolution with limited knowledge. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING) Conference*, pages 869 – 875.

[Rocha, 1997] Rocha, M. (1997). Supporting anaphor resolution in dialogues with a corpus-based probabilistic model. In *The Proceedings of the Workshop on Operational Factors in Practical Robust Anaphora Resolution in Unrestricted Text (ACL)*, pages 54–61.

[Rocha, 2004] Rocha, M. (2004). The classification of anaphoric demonstratives into processing strategies for automatic resolution: A corpus-based approach. In *Proceedings of the 5th Discourse, Anaphor Resolution Colloquium (DAARC)*, pages 139 – 144.

[Sinclair, 1992] Sinclair, J. (1992). Priorities in discourse analysis. In Coulthard, M., editor, *Advances in Discourse Analysis*, pages 79–88. Routledge.

[Soon et al., 2001] Soon, W. M., Ng, H. T., and Lim, D. C. Y. (2001). A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4):521–544.

[Strube, 1998] Strube, M. (1998). Never look back: An alternative to centering. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics*, pages 1251–1257.

[Strube and Hahn, 1999] Strube, M. and Hahn, U. (1999). Functional centering: Grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344.

[Stuckardt, 2001] Stuckardt, R. (2001). Design and enhanced evaluation of a robust anaphor resolution system. *Computational Linguistics*, 27(4):479–506.

[Webber, 1991] Webber, B. (1991). Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes*, 6(2):107–135.