Automatic Resolution of Anaphora in English

Michel Denber

Eastman Kodak Co.
Imaging Science Division
June 30, 1998

Table of Contents

Acknowledgments

Thanks go to Jean Banker of the Strong Musesum in Rochester, NY for providing extensive real-world image caption data for use in this project, as well as to Gian Carlo Cervone of the George Eastman House for data he provided. Both of them expressed an interest in the subject and expressed their willingness to provide further assistence. I also appreciate the help of the ISTL lab members who volunteered to help create the image caption set actually used for testing. They include Steve Etz, Bob Gray, Rajiv Mehrotra, Marty Petrella, George Sotak, Pete Stubler, and Amy Zhu.

The staff of the ISD Computing Resources Group were invaluable in helping get various software components installed and running. Randee Tengi at Princeton University generously provided files not normally available for use in reconfiguring WordNet. Finally, I greatly appreciate the assistance of Rajiv Mehrotra and Bob Gray who reviewed the draft of this report and provided many excellent comments and suggestions.

Abstract

An algorithm is described for the automatic resolution of inter-sentential anaphoric references in English sentences. Anaphora are backward references to nouns which have previously been mentioned in the input text stream, such as *his* in the sentence *John started his car*.

The method is based on constraint matching and makes use of both syntactic and morphological information about the sentence structure and real-world knowledge about the semantics of the sentence elements gained from several different sources to generate features for use in matching.

The ability to perform anaphora resolution is important in multimedia image retrieval for improving the completeness of image query and caption understanding, leading to better matches in retrieved images. For example, given the query

A man and a woman.

He is standing, she is sitting.

a simple word-matching retrieval algorithm would return images containing all combinations of men and women both standing and sitting. However, if the anaphoric referents to *he* and *she* could be resolved, the accuracy of the results returned could be improved by knowing who was doing what. As the number of images stored electronically with captions continues to increase, this ability will become increasingly important.

1 Introduction

1.1 Definitions

The linguistic term *anaphora* is an exotic name that refers to a common phenomenon. Like Molière's Monsieur Jourdain who discovered with astonishment that he had been speaking in prose without even realizing it [Moli], everyone uses anaphora on a daily basis whether they know it or not. The etymology of the term goes back to Ancient Greek with *anaphora* (αναφορα) a compound word consisting of the separate words ανα- *back, upstream, back in an upward direction* and φορα - *the act of carrying* and denoting *the act of carrying back upstream* [thanks to R. Mitkov for this definition]. Anaphora are forms of reference in written or spoken speech in which a word, most commonly a pronoun is used in place of a previously mentioned item (most often a noun or noun phrase) where both refer to the same entity. For example, in the sentence,

Kathy always took her camera with her.

her refers to Kathy (in both instances).

Much as is the case with visual interpretation of images, humans normally perform anaphoric resolution completely unconsciously and usually with high accuracy. However, there are cases in which the referent is ambiguous. For example:

While Jack was talking, Joe finished his beer.

Here *his* may refer to either *Jack* or *Joe*, largely depending on the context of the situation.

The automatic resolution of anaphora by computer is arguably one of the most difficult problems in natural language processing because it depends heavily on both linguistic knowledge (how sentences are constructed) and world knowledge ("common sense" facts about how the world works).

This paper presents an algorithm for automatically resolving two forms of two-sentence intersentential pronominal anaphora, i.e. ones in which the anaphoric co-referents appear in the sentence directly preceding the one containing the anaphora themselves, in American English. Although the operation of the algorithm is described in detail, the primary point here is to serve as a sort of roadmap of the entire topic so that someone revisiting this issue in the future can have some idea of the problems that might be encountered and the extent of the difficulties involved as well as some history of what has come before.

1.2 Motivation

Anaphora resolution is obviously useful in a speech or text understanding system in order to help the machine understand what is going on. Indeed, it can be said that an understanding of the co-referents in anaphora *is* a good part of what is going on in a realistic spoken or written discourse or text.

Similarly, anaphora resolution can assist the process of image retrieval by natural language query. This application resembles speech processing in the need for interactive response. It would be of interest to customers such as stock photography houses interested in offering users natural language image query that works better than simple keyword search, museums wanting to improve access by researchers or the general public to their collections, as well as individual consumers captioning their own snapshots for digital storage. For example, the query

A lake with the sun reflected in it.

was posed to the search engine at the Publisher's Depot Web site. Of the 12 images returned on the first page, only three were relevant to the query. When the query was rephrased as

A lake with the sun reflected in ***lake****.*

the number of relevant images increased from three to eight.

Consider as well the example caption,

A picture of a bride and groom.

He is putting the ring on her finger.

A search engine which performs straight keyword matching will return all images that contain the words *bride, groom, ring,* and *finger*. While this combination will most likely return an image with the desired content (assuming one exists in the database) it will also return other unrelated images, such as

The reverend points his ***finger*** *towards the* ***ring*** *on the altar between the* ***bride*** *and* ***groom****.*

which contains all of the keywords but describes a rather different image.

If on the other hand, one could resolve the anaphora *he* and *her*, the original caption could be restated as

Groom is putting the ring on bride's finger.

In this form, a semantic association can be made associating *ring* with *groom* and *finger* with *bride*. This would help improve the accuracy of retrieved images. Research has shown that anaphora resolution does in fact improve both recall and precision performance in largescale database queries [Pirk]. While this may not make much difference if only a dozen or so images are found, it starts to become more important as the size of the database grows. For example, users may be satisfied even if 50% of 12 images are irrelevant, but may be considerably less satisfied if they have to search through several hundred images looking for the relevant 50%.

1.3 Prior art

To my knowledge, this is the first instance of work specifically on anaphora resolution within Kodak. It was intended to complement work done by Archna Bhandari for performing natural language query of image captions, and which is based on Fillmore's case grammar [Fill] but does not specifically address anaphora resolution. The program presented here does not attempt to match queries with captions, it only associates anaphora with their antecedants. The "de-anaphorized" text can then be used as input to the query-caption matcher to (hopefully) improve the accuracy of the match as described in the preceding section.

Work on anaphora resolution in the open literature tends to fall into three domains: *artificial intelligence* (as a specialty of computer science, including computational linguistics and natural language processing), *classical linguistics* (as distinguished from computational linguistics), and *cognitive psychology* [Lust 82, Sanf]. Psychologists tend to be interested in this topic because of what they can learn about how the brain processes language. Linguists are interested simply because anaphora resolution is a classical problem in the field. For our purposes we are primarily interested in the AI/computational linguistics approach.

The various approaches to anaphora resolution differ largely in the depth of linguistic analysis being performed, that is the extent to which the system tries to understand the contents of the input text. At one extreme is the work of Kennedy and Boguraev [Kenn] which is purely morpho-syntactical in nature. This has the advantage of being fast and relatively easy to implement. A middle ground is found in Lappin and Leass' RAP program [Lapp] which is also based on surface processing but employs a more extensive syntactic analysis. At the other extreme are discourse-understanding systems [Alla] which attempt to build a model of what information is being converyed in the input. These can be quite complex and use a variety of techniques including various corpora, statistical techniques [Daga], fuzzy logic and neural networks [Conn] most commonly associated with artificial intelligence work. Some systems, notably by Mitkov [Mitk94] try to gain advantage by combining different approaches. Two key components widely found in most of the existing work are the use of a separate parser or tagger (which performs a limited version of parsing) and the use of an external lexicon most commonly to provide basic lexical information but sometimes also to derive semantic clues. Although a variety of taggers are in use, the *WordNet* system from Princeton University is by far the most popular lexicon in the field.

A number of linguistic ideas are often applied to computational anaphora resolution, including centering theory [Kehl 93], the notion of c-commanding [Alla], and discourse representation theory (DRT) [Fisc].

Most of applications targeted in the existing work deal with natural language processing for question-answering systems, machine translation, and data mining. In fact, no references were found for systems specifically addressing image query and caption matching. However, interest in anaphora resolution is currently quite high and several conferences devoted entirely to anaphora resolution have been held in the last two years (cited extensively in the references).

The following major companies are known to have interests in anaphora resolution: Apple [Kenn], BBN, IBM, Kurzweil, Microsoft (which has a group of more than 30 researchers devoted to natural language), NEC, Quarterdeck, Sony, SRI [Kehl], and Xerox [Dalr]. They are motivated by interests in text and/or speech understanding for applications including multimedia image retrieval, data mining, document understanding, and machine translation of natural languages. In addition, numerous small companies are active in this area for the same reasons.

Very frequently cited works in the literature include those by Hobbs [78], Lappin and Leass, Luperfoy, and Sidner.

It is interesting to note that as of this writing there are no U.S. patents issued that contain the string *anaphor**. The patent situation is treated separately in Section 7.

1.4 Types of anaphora

Although there is no unanimous agreement in the literature as to how they should be categorized, the (more extensive than most) approach taken here is to define six main types of anaphora, plus one "catchall" class. The succinctness of the taxonomy belies the complexity of the problem. The classes are:

1.4.1 Pronominal

Pronominal anaphora are the most commonly encountered in general usage. This category includes three subclasses:

- Personal: *he, she, it, his, hers, its, they, them, their, theirs*
- Demonstrative: *this, that, these, those, others*
- Reflexive: *himself, herself, itself, themselves*

Within these groups, the personal pronouns constitute the great majority of anaphora in general texts. The demonstrative pronouns are quite variable. They can be anaphoric as in,

*The car wouldn't start because of **this**.*

but are probably more often seen as cataphoric (see below) as in

***This** is a ruby laser.*

1.4.2 Quantifier

One (two, three…n), each, some, many, most, all, every, none, a little, a few, a lot (of them)

This form can occur as both subject and object:

***One** was green.*
*She chose the blue **one**.*

1.4.3 Ordinal

(*The) first, second, etc., former, latter*

These forms are anaphoric depending on context. For example, in

*His horse came in **first**.*

first is not an anaphor while in *The **first** was easy.*, it is. However, this appears to be encountered relatively less often than the equivalent *The first **one** was easy.* (which becomes a *one*-anaphora).

1.4.4 Question

Also called *wh-anaphora*, includes the pronouns

who, what, which, where, when

as in

*Even before the butler confessed, everyone knew **who** did it.*

*Main Street is **where** the action is.*

1.4.5 Event anaphora

Anaphora need not have only a single noun as a referent. This is distinct from the common generalization of the referent to a complete noun phrase. Consider for example,

In 1969 America landed a man on the moon.

***It** was NASA's shining moment.*

The first sentence contains four nouns (*1969, America, man, moon*) but *it* refers to the entire *event "the moon landing"* or *"landing a man on the moon"* rather than either any single one of them or to the NP *a man*. These types of anaphora require a deeper understanding of the semantics of the sentences involved to resolve properly than is available simply from a syntactic and lexical analysis and are therefore also beyond the scope of this paper. To the extent that the referent is really a paraphrase of the actual antecendent text (as in this example), this problem also borders on ellipsis (see below).

1.4.6 Pleonastic *it*

This is an interesting case of a pronoun which does not refer to anything. (*It* is raining. *It* is summer. *It* seems that… *It* just so happens that… *It's* late… *It's* 3 'o'clock., *It's* about time.) This form is also called *non-referential* or *expletive it* and is considered in further detail in Section 2. In the literature, pleonastic pronouns are commonly not considered to be anaphora (precisely on the grounds that they *don't* have referents), but it seems that they need to be recognized in a system that handles anaphora, if only to know not to look for a referent for them. Therefore, they are included as a separate category here. Actually, one could even make the case that pleonasms do have referents of a sort, such as "*the time*" or "*the weather*", however, this point is debateable.

1.4.7 Special cases

These forms tend toward the idiomatic and there is not always even agreement as to their status as anaphora. Two common examples include

Do so, and *respectively.*

There are also *temporal anaphora* (as distinct from the time pleonasms given above) which are important in discourse representation. These are touched on in section 2.7.2.2.

In addition, anaphora may be *intrasentential*, where the referent is in the same sentence as the anaphor (*Jack tried to start his car.*) or *intersentential* where the anaphor refers to a noun in the previous sentence (*Jack got in his car. He tried to start it.*). Anaphora may even refer back more than one sentence in a discourse, although the more widely separated the elements are, the less likely it becomes that they may be co-referents. It is expected that this situation is also less likely to be encountered in a query/caption system where the individual textual units tend to be short.

1.5 Related constructions

There are other types of linguistic constructions that appear similar to anaphora but are not strictly classified as such. For example, a sentence may contain a forward reference, as in

When he replaced the water pump, Joe scraped his knuckles

where *he* refers to the subsequent mention of *Joe*. Technically, this type of construction is called *cataphora*, although in the linguistics literature anaphora is sometimes presumed to include cataphora (however, because they involve different analysis strategies, we will not do so here). In this example both pronouns refer to *Joe* but *he* is cataphoric and *his* is anaphoric. Both anaphora and cataphora are forms of *endophora* (within-context references). For completeness, it should be mentioned that there are also *exophora*, which are “universal” or commonly understood referents (possibly in a particular context assumed to be shared between the speaker or author and the audience) such as *the sun*, *the PC*, or *the President*. One copier company attempted to exophorize itself by adopting the slogan “the Document Company”.

Next, *ellipsis* is a linguistic form which resembles anaphora except that here it is the *absence* of something that refers to something else. For example, in

Uneven stitches will result if the upper thread tension does not match the lower.

there is an elided *thread tension* at the end of the sentence. This example can be converted to an anaphor by adding *one*, as in

Uneven stitches will result if the upper thread tension does not match the lower ***one****.*

One quite common example of anaphora and cataphora in ellipses involves *this* as in

This is a picture of uncle Joe.

where *This is* contains an elided *picture (This* ***picture*** *is a picture of uncle Joe.)*

In any case, elliptical constructions are quite difficult to recognize and beyond the scope of this paper.

Finally, for completeness we examine forms like

It's beginning to feel a lot like Christmas.

Which at first glance appear to be pleonastic since *it* generally has no direct antecedant. However, here *it* really does appear to refer to something, if only an abstraction or particular seasonal gestalt. This is more of an "elliptical pleonasm" than a true anaphor.

With this understanding of the anaphora resolution problem, we now examine a method of solving it by computer in the next section, with a specific focus on intersentential anaphora for use in image query and caption analysis (implying a limited amount of text)..

2 The algorithm

2.1 Design

The algorithm implemented here is called *AM*, for *Anaphora Matcher*. It is designed to handle inter-sentential anaphora over a two sentence context. Currently it can handle two types of anaphora: *personal pronouns* and *pleonastic its*. These were selected since the former is by far the most common category, while the latter could cause considerable confusion in sentence meaning if it was attempted to find resolvants for it. Also since *it* appears in both categories, no treatment of personal pronouns is complete without also considering pleonastic *its*. A two-sentence "window" is reasonable for anaphora resolution in the domain of image queries and captions since it is expected that few of these types of expressions would exceed two sentences in length. The algorithm could be extended beyond two sentences by keeping information from previous sentences in context (the "history list" idea [Alla]) or simply by considering the input two sentences at a time. In any case, most anaphora do not refer back more than one sentence. Multi-sentence intersentential anaphora resolution also has been studied using centering theory [Kehl 93] however, this type of approach increases the complexity of the solution and is beyond out scope.

AM is based on *constraint-matching*, using both syntactic information about the sentence as well as real-world semantic knowledge obtained from several sources. The idea is based on the way humans perform at least some forms of anaphora resolution. For example, in the sentence

Joe and Melissa walked in the room and she sat down.

the most likely interpretation is that *she* refers to *Melissa* and cannot refer to *Joe* because there would be a gender constraint mismatch in that case. Some algorithms attempt to solve the anaphora problem purely syntactically [Lapp]. While there may be good reasons for doing this, it greatly increases the difficulty of an already difficult problem and there is no reason to be so restrictive if the computational resources are available to perform more in-depth analysis. By using all of the information available, we can generate better matches. AM uses three different external sources of information in its processing:

1. AM makes use of a commercially available natural language tagger called *engcg* (for *Eng*lish *C*onstraint *G*rammar) based on the theory of constraint grammar [Karl].. engcg is one of a number of so-called "super taggers" now available and classifies each word in an English sentence according to its part of speech (noun, verb, etc.), morphological features (singular, plural, etc.), and syntactic role (subject, object, etc.). It is not strictly a parser in the sense that it does not partition a sentence into the traditional noun phrase/verb phrase groupings, however, it does provide important information for use in anaphora resolution.

2. AM also uses the freely available *WordNet* lexical database system [Fell] to obtain real-world knowledge about the words in the input sentences. WordNet is basically a

combination dictionary/thesaurus organized hierarchically. It provides not only definitions and synonyms, but part-of and kind-of relationships between words as well as frequency data which can be useful for determining the more likely alternative in cases of ambiguity. WordNet is quite popular in current natural language processing research.

3. AM uses a set of three lists of proper names from the MIT Media Lab as part of a knowledge representation database package there [Haas]. These categorize proper names as either male names, female names, or place names.

2.2 Detailed procedure

The algorithm takes as its input two English sentences provided as a single string and returns as output a list of any of the following: the co-referents in the first sentence of any anaphora in the second sentence, "none" if no anaphora were present, or "pleonastic" if an *it* was recognized as such. Steps 1 and 2 below pre-process the input. Steps 3 through 7 deal with assigning constraints to the possible referents. Steps 8 and 9 apply constraints to the anaphora, and step 10 is the actual matching operation. It proceeds as follows:

1. The two-sentence input is provided as a single string. The first step is to divide this string into two sentences. This is done by searching the string for a sentence delimiter in the form of ".", "!", or "?". Everything up to that point is considered to be the first sentence (S1) and the rest of the string is considered to be the second sentence (S2).

2. The first input sentence (S1) is passed to engcg for syntactic analysis. A list is made of all nouns in the sentence using the part of speech tags returned by engcg. This becomes the possible solution set for any anaphora in S2. Since AM is specifically designed to handle intersentential anaphora, S2 is not searched for resolvants (although some might exist there).

3. For each noun found in S1, it is examined to see if it is a noun-noun modifier (i.e. a noun acting as an adjective as in *the **log** cabin, the **wood** elephant*) by checking its part of speech. This is done by making a list of the part of speech of each word in the input and searching for the pattern $\mathbf{N_a^+ N}$ (a string of two or more consecutive nouns). All of the $\mathbf{N_a}$ nouns are removed from consideration as they are considered ineligible as anaphoric antecedents.

4. Each remaining noun is tagged for number (singular or plural) and syntactic function (subject or object), again using information provided by engcg.

5. Each noun is tested for capitalization and one of step 6 or 7 is executed as a result.

6. If a noun is capitalized, it is assumed to be proper and an attempt is made to determine if it is a place name or a personal name by looking for it in a set of name and place lists. For personal names, an attempt is made to tag them with the correct gender using gender-typed name lists following the procedure described in detail in section 2.4 below. This information is added only if a name exists uniquely in one list and not the others, i.e. rather than risk guessing wrong, ambiguous names are

simply left untagged. Note that the test "capitalized implies proper noun, else common noun" is only a heuristic. One can easily imagine circumstances in which it would not hold. However, in the absence of any better way to distinguish proper nouns (which is a difficult problem in its own right) this will have to do.

7. If a noun is not capitalized, it is assumed to be common and WordNet is searched for its *hypernyms* (*kind-of* relationships). This yields information concerning the noun's classification as animate or inanimate as well as possibly its gender. If such information is found, the appropriate tags are added to the noun's constraint list. The animate/inanimate tag is set by searching WordNet for the word *creature* or *object*. The gender tag is set by searching for *male* or *female*. Nouns with the object tag also receive the neuter gender tag.

8. The second sentence (S2) in the input is sent to engcg for analysis. In this sentence, all of the second- and third-person pronouns are collected as the anaphora which must be resolved. Each of the pronouns which AM knows about has an associated list of constraints defined within the program (for example, *he* has the constraints *masculine* and *singular*).

9. Using the results of engcg, syntactic constraints are added to each pronoun specifying their role in the sentence as subject or object (assuming engcg returned that information which is not always the case). One could imagine using more of the case grammar tags and applying various weights to them to provide dynamic constraints as is done by Bhandari. Although this approach has in fact been used by some authors [Carb][Fisc], it appeared to introduce additional complexity with an uncertain payoff and was not pursued.

10. The list of pronouns as collected in step 8 is examined. Any *it* is first tested for being pleonastic. For any other pronoun, as well as non-pleonatic *its*, its constraint list is matched against that of each noun in S1. The noun which matches the greatest number of constraints is considered to be the antecedent of the pronoun. In the event of a tie, the noun which matches in the subject/object constraint is judged the winner. If this does not break the tie, or in the case of a tie where that the noun or pronoun was not assigned a syntactic category by engcg, the first noun encountered in S1 is used. Syntactic role agreement was used as a tie-breaker rather than earlier in the matching process because it was deemed to be less reliable as an indicator of co-reference than the other constraints used. Matching a noun does not remove it from the list of candidates for matching subsequent pronouns, since it is common for more than one anaphor to refer to the same noun.

The aim of this algorithm is to maximize the amount of information available to make a decision while minimizing the consequences of missing or conflicting information. For example, if a proper name is encountered which could be either a male or female name, then the gender constraint is simply omitted. This reduces the amount of information available but does not affect the operation of the matching procedure. In this type of situation, a human observer would have the same problem.

2.3 Number agreement

Number agreement refers to the grammatical constraint found in many natural languages including English requiring that certain pronouns (such as *he*, *she*, and *it*) only be used with singular nouns and others (*they*) only with plural nouns. English pronouns must also agree in number with verbs although this fact is not exploited here. The number constraint is important because it is always known for pronouns and is usually easy to determine for antecedents. The engcg parser is quite good (but not perfect) at assigning the proper number classification to nouns. For nouns that are ambiguous in number, such as *deer* or *fish*, engcg returns the tag "SG/PL". However, in these cases a human would also be unable to distinguish them; the computer can do no better. Of course, the use of number agreement as a constraint in AM assumes that the input will follow the grammatical rules of English.

The remainder of this section discusses number agreement problems that are not addressed by AM. One common problem involves the process of set formation. For example in the sentence pair

The man and the woman sat down to eat.

***They** ordered steak.*

they is a plural pronoun but both possible antecedents (*man* and *woman*) are singular. The correct answer is "man and woman". The same problem arises with *or*:

*Tom had to choose vi **or** emacs but he hated **them** both.*

It can also happen without any conjunction at all:

Joe took the boys to the ball game.

***They** had a good time.*

Here *they* refers to both *Joe* and *boys*, at least in the most common interpretation (collective reading).

Sometimes number agreement fails. This happens with certain nouns that look plural but act singular, as in

*Nancy took the kids to the **movies**.*

*They argued over which **one** to see.*

Movies is plural, but it is the correct antecedent to *one*, which is singular. Other members of this class include *the races*, *the fights,* and *the falls.*. One possible solution is to regard *movies* and "***the** movies"* as two different entities with the latter being treated as a singular construction.

In addition, sometimes people will deliberately violate number agreement by using *they* in order to avoid using a gender-typing pronoun as in

*When the **user** starts the program, **they** will be asked for a password.*

In this case, *they* is used to replace the phrase *he or she*. Although this is not considered good usage, it does occur in practice.

2.4 Gender agreement

Gender agreement refers to the linguistic requirement that a pronoun and co-referring noun match in grammatical gender. In English this is restricted to proper names, where *he* and *she* are used to refer to males and females respectively and *it* is reserved for inanimate objects or animals of undetermined gender.

Therefore, proper names can be useful clues as potential antecedents in that they often contain constraints on gender and animacy. Unfortunately, they also present a surprising number of problems because of potential ambiguities. Proper names can be recognized in English by virtue of being capitalized. This is the extent to which AM deal with proper names. Unfortunately, the first word of a sentence is normally always capitalized, so if that word is a noun it can't be identified as proper that way. This same problem extends to some words such as *Count*, which can be a title or a verb when it occurs at the start of a sentence. Also, older legacy databases are often found to have been entered entirely in upper-case characters.

This is important because WordNet contains no proper names and one wants to avoid sending it one since it is not uncommon for proper names also to be common nouns which could return misleading information. For example, in WordNet *Bob* returns data about hairstyles, and *Jack* returns data about devices for lifting cars, neither of which have the desired gender attributes for those names. There is also considerable overlap between girls' names and flower names (*daisy, heather, ivy, rose*, etc.) for example. Ultimately of course, names can be anything as in Frank Zappa's daughter *Moon Unit*, the artist formerly known as Prince, various nicknames, etc.

Some proper names can be ambiguous in gender (e.g.; *Chris*, *Pat, Tracy*). In addition, some names differ in gender in different languages. This needs to be considered since foreign names can easily appear in English sentences. For example, *Jan* is a boy's name in Polish and a girl's name in English; *Jean* is a girl's name in English and a boy's name in French. There is further ambiguity between person names and other proper nouns such as place names or organization names. For example, *Troy* can be both a boy's first name and a city (in fact one of several different cities). Other names may be ambiguous with non-place proper names. For example, *April* is both a girl's name and a month of the year. Table 1 summarizes the possible intersections in the name lists used by AM.

	male names	female names	place names
male names	2947	374	159
female names		5001	97
place names			4788

Table 1 : Proper name intersections

When it is not possible to unambiguously resolve a proper name for gender, no gender tag is attached to it.

Proper names also have a scoping problem which causes referent ambiguity. For example, in

Sol Hurok was an impressario.

He brought the Bolshoi to America.

one would like to resolve *he* with *"Sol Hurok"* rather than *Sol* or *Hurok*. A heuristic for this might be to check the word following any proper name. If it is either a proper name or a proper noun of undetermined type, group it with the preceding name as two-name compound noun. This would also solve the related problem of multiple first names such as *Joe Bob* or *Emma Sue*.

As with many features in natural language, there are some anomalies relating to gender in English. It is a custom to refer to ships and large boats as *she*. No attempt is made to identify such instances here as it is a difficult problem and expected to occur relatively infrequently. For completeness we add that sometimes natural features or processes are anthropomorphized and these tend to be identified as *he* (e.g. *Ol' man River, **he** jus' keep rollin'*. or *When the print manager sees a new job, **he** queues it up.*). These forms are also not addressed here.

2.5 Syntactic agreement

Empirical evidence suggests that anaphora commonly match their referent in their syntactic roles [Alla]. In addition to the lexical and semantic constraints mentioned so far, AM makes use of one syntactic constraint. It attempts to match the syntactic role of the anaphora (as subject or object only) with that of their possible co-referents. This helps match examples such as

The sun is setting over a lake.

***It** is reflected in **it**.*

Here, the first *it* (the subject) refers to *sun* which is also the subject of its sentence and the second *it* (the object) refers to *lake* which is correspondingly the object of the first sentence.

Prepositional phrase ambiguity is an interesting problem in relation to syntactic role constraints. Consider the following example and two sucessors:

At this point, EMACS is waiting for a command.

***It** is prepared to see if the variable keys are true.*

***It** should be entered in lower-case.*

In the first anaphoric sentence, *it* refers to *EMACS* (the subject) while in the second *it* refers to *command* (the object). However, in both cases *it* is the syntactic subject. A constraint match based on this feature would match both *it's* with EMACS. Argument could be made here that what is really needed is the *deep structure* subject (in the transformational grammar sense [Chom]). If we rephrase the second sentence as

***Someone** should enter it in lower-case.*

(removing the surface passivization construction) then *it* becomes the object and would correctly match in role with *command*. Unfortunately, to identify this would require the implementation of a transformational grammar on top of the existing anaphora matcher and this is beyond the scope of the current work.

2.6 Animacy agreement

For the pronouns *he* and *she* in English, it is possible to apply a constraint for animacy *vs.* inanimacy. This is distinct from the constraint on gender and is useful in identifying animals whose gender is not otherwise specified, as in the caption pair

A picture of a dog carrying a stick.

He is running in circles.

Here it is possible to distinguish the co-referent of *he* as *dog* instead of *stick* on the basis that *dog* is tagged as animate in WordNet but *stick* is not. This is implemented in AM by searching the WordNet hypernym hierarchy of the noun for the keyword *creature*.

Proper names cause problems with animacy constraints. For example, *George Washington* is animate but *the George Washington Bridge* is not. This assumes that *George Washington Bridge* can even be identified as one object, which is not always the case as mentioned in section 2.4. AM does not handle these forms.

2.7 Pleonastic *it*

2.7.1 Definition

The term *pleonastic* (from the Greek *pleion*, "plus" or as used here, "more") is the adjectival form of *pleonasm*. A pleonasm is a linguistic form where more words are used than needed to convey the intended meaning, i.e. a redundancy. The pronoun *it* is said to be pleonastic when it is used in a context where it has no referent. Lewis Carroll described the phenomenon thusly:

> "I proceed. 'Edwin and Morcar, the earls of Mercia and Northumbria, declared for him; and even Stigand, the patriotic archbishop of Canterbury, found it advisable –'"
>
> "Found *what?*" said the Duck.
>
> "Found *it*," the Mouse replied rather crossly: "of course you know what 'it' means."
>
> "I know what 'it' means well enough, when *I* find a thing," said the Duck: "it's generally a frog or a worm. The question is, what did the archbishop find?" [Carr]

It is interesting to note that despite an extensive treatment of anaphora in the computational linguistics literature (see section 8), only two instances have been found where the problem of pleonastic references is addressed [Kenn][Lapp], and one of them [Kenn] only appears to reference the other. In addition, Ruslan Mitkov, a noted expert on the topic of anaphora resolution, agrees [personal communication] that this problem has been essentially ignored. Although this does not necessarily *prove* that no such work exists, it calls to mind the cartoon in Garey and Johnson's classic book on NP-completenesss [Gare]:

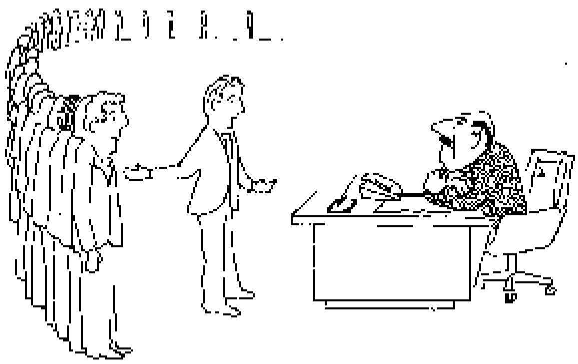

Figure 1 : *I can't find an efficient algorithm, but neither can all these famous people.*

As Garey and Johnson say, "At the very least, this would inform your boss that it would do no good to fire you and hire another expert on algorithms."

In fact, it seems quite surprising that in anaphora resolution work in general, it is considered standard practice to remove pleonasms by hand from the database being used for testing. This is justified by the claim that pleonasms are not truly anaphora since they have no referent. While this is true, it disregards the fact that the computer has no way of knowing *a priori* whether an *it* it encounters is pleonastic or not. Therefore it would appear necessary to give a reasonable accounting of pleonasms in any work on anaphora resolution.

AM has a module for performing pleonastic resolutions which are divided into two main categories, *state references* and *passive references* as described in the following sections.

2.7.2 State references

State references branch into two main forms: *temporal* and *meteorological*. Pleonastic-*it* is commonly involved in descriptions of the weather, for example, *it's sunny*, *it's raining*, etc. and in statements about time as in *It's three o'clock, It's time to go,* etc..

2.7.2.1 Meteorological pleonasms

AM tests for meteorological anaphoric references by checking for three different syntactic patterns:

1. ***it <be>* ADV* PCP1** where the verb *be* may appear in any inflected form, the * indicates "zero or more occurances of" in customary finite state expression form, and

the PCP1 (present progressive) verb appears in a list of 12 common meteorological verbs (*snowing, raining, drizzling,* etc.). This matches forms such as *It is raining.*, and *It is hardly raining*.

2. ***it <be>* ADJ** where the verb *be* may appear in any inflected form and the ADJ appears in a list of 44 common meteorological adjectives (*cloudy, sunny, windy,* etc.). (e.g. *It is cloudy*.)

3. ***it <be>* N** where the verb *be* may appear in any inflected form and the N is one of five season nouns (*spring, summer, autumn, fall, winter*).

It is not unreasonable to believe that such forms might appear in image captions such as:

This was taken just before it started raining.

This was when it was summer.

2.7.2.2 Temporal anaphora

In the literature [Nelk][Part], the term *temporal anaphora* is used to cover words such as *often, before, after,* and *when* in situations such as

***Before** taking off on a jump, Katarina Witt **always** crosses her arms.*

This type of anaphora has not been studied extensively and is not addressed here since it is more important for discourse understanding than pronoun resolution. Instead, we examine pleonastic time-expressions such as

1. *It's three o'clock.*
2. *It's time to go.*
3. *It's past time to go.*
4. *It's about time.*
5. *It's late.*

The first form (containing a definite time) could be handled with a pattern of the form

***it <be>* ADV* NUM**

where NUM matches a time expression (which is parsed as a unit by engcg). Forms 2 through 4 (containing relative time references) could be handled with

it <be>* PREP* *time

Form 5 will match

***It <be>* ADV* (*early* | *late*)**

Interestingly, there do not seem to be any other members of the set in form 5 besides *early* and *late*.

None of these patterns is currently implemented in AM.

2.7.3 Passive construction

The other common form of pleonastic-*it* typically involves passive forms such as *it seems that…, it can be seen that…, it is clear that…*etc., as demonstrated by the Mouse's speech in the previous section. Lappin and Leass [Lappi] describe three syntactic patterns in which these modal adjectives can appear in pleonastic usage. They are incorporated into the design of AM in a slightly revised and generalized form as follows:

it* <be> modadj *that

***it* <be> modadj [*for* N] *to* V**

N Vcog *it* modadj [*for* N] *to* V

In addition, we add a fourth pattern:

***it* <be> modadj PRON**

This matches expressions such as *It seems he left…*

Lappin and Leass also define a set of 15 modal adjectives (including their negations, comparatives, and superlatives) that may appear in these patterns: *necessary, good, economical, possible, useful, easy, certain, advisable, desirable, likely, convenient, difficult, important, sufficient,* and *legal*.

Of course, the problem is that this list is far from exhaustive (one obvious missing adjective is *obvious*). It is here that WordNet can be particularly useful. Adjectives in WordNet can appear with an *attributes* property. It is proposed that an adjective qualifies as modal if it has an *attributes* property and those attributes include one or more of *state, condition, quality,* and *quantity*, but not *color* or *physical*. In addition, any adjective that ends in *–ble* qualifies as modal in this context. Of course, many adjectives that end in *–ble* are not modal (*variable, edible, adjustable,* etc) but none of these may appear in the context of the above patterns in semantically well-formed sentences, for example,

It is* ***variable* *that the man left.*

vs.

It is ***understandable*** *that the man left.*

Lappin and Leass also define seven cognitive verbs (*recommend, think, believe, know, anticipate, assume, expect*) that may be used in the following pattern:

it* <be>* Vcog+past *that

This matches expressions such as *It seems that…, It is thought that,* etc. Again, there are many more cognitive verbs than these (*feel, note, observe*, etc.). Fortunately, engcg tags these verbs with the feature <Vcog> so it is not necessary to maintain an explicit list.

Other linguistic elements may be pleonastic in certain contexts. For example, it can be argued that *else* is plenostic [Culi] in

She didn't want lasagna, she wanted something ***else****.*

However, *it* is by far the most common form of pleonasm in common usage. None of the patterns in this section are included in the current implementation of AM.

3 Results

3.1 Test data

AM was tested on a sample database of 82 two-sentence image captions containing a total of 48 anaphora. No attempt was made to exclude captions which did not contain anaphora in order to more fully test the operation of the program. The captions were collected from various members of the Imaging Science Technology Laboratory who were asked to view random images from the JBJL image database and create captions for them as if they were snapshots they had taken themselves and they wanted to annotate them so that they could remember what they were about in the future. The only restriction was that the captions should be two sentences long, since the program being tested was designed to handle intersentential anaphora. Although the captioners were not specifically requested to use anaphora in their captions, they sometimes did so. This demonstrates that anaphora may actually occur in real captions. The author then reviewed the captions and hand-generated ground-truth data for each anaphor found, explicitly listing its co-referent or "pleonastic" as appropriate. There was one instance where it was deemed impossible to unambiguously determine the antecedent of an anaphor; this instance was labeled "unknown". Table 2 below indicates the breakdown of the 48 instances of anaphora by type according to the taxonomy of section 1.

Type	**pronoun**	**quant.**	**ordinal**	**pleonas.**	**ambig.**	**none**
Number	41	0	0	6	1	51

Table 2 : Test data anaphora distribution

The AM algorithm was run on this test data. Only the pronominal and pleonastic anaphora were included in the scoring since instantiation anaphora were not implemented. There were six demonstrative pronouns (*this*) which were also excluded from the results since these forms are also unimplemented. Also, eight captions containing a total of 13 anaphora had to be omitted because they were not in the proper intersentential form.

Finally, the lone ambiguous anaphor was omitted since there was no reasonable way of deciding if the program's answer was correct or not. An alternative approach would be to include ambiguous references and consider them correct if the program chose one of the reasonable referents. Or it could be argued that in order to be truly "correct", the program should return *all* of the possibilities in the event of an ambiguity. In actual practice, the program will always make a match if at least one possible antecedent exists, unless the referent is found to be pleonastic. In this particular test, the program did choose one of the reasonable alternatives. However, dealing with ambiguous anaphora simply adds complication on top of complexity and was deemed to be beyond the scope of the present work, therefore the ambiguous case was omitted from the results.

Of the remaining 24 anaphora, AM identified 10 correctly, or 42%.

The computational complexity of this algorithm is time O(n * m) where n is the number of nouns in the first input sentence and m is the number of pronouns to be resolved in the second sentence. Typical running time of the algorithm for a two-sentence input with five nouns and two pronouns is approximately 55 seconds in real-time, although this figure is not very useful, since all of the interprocess communication between AM and both WordNet and engcg occurs via networked disk files which incur lengthy delays and the proper name lookup is conducted from within Lisp as a simple linear file search, which itself is exceedingly slow. The actual CPU time spent in performing the resolution itself is well under 1 second on a Sun Sparc 5 workstation.

3.2 Error analysis

Three of the 14 errors AM made, or 21%, were due to faulty information received from engcg. In two cases it reported that the noun *Molly* was being used as an adjective in the construction

Molly, age 8, …

This automatically excluded *Molly* (the correct referent) from further consideration. In the third case, engcg incorrectly identified *Sue* as a verb in the sentence

Sue in a maternity nightgown.

From a purely syntactic point of view the verb reading is legal, however it does bring to mind a rather odd courtroom scenario when viewed in that context. This is one of many examples of sentence fragment problems which are discussed more in section 4.1.1. One solution in this instance might be to have a rule to the effect that when the first word in the sentence is reported by engcg to be a verb, check in the name lists to see if it can also be a name. If so, then prefer the name form. The reasoning is that captions and queries are probably more likely to be descriptive than imperative. Other similar examples include *Bill in a tuxedo., Bob in the pool.,* and *Carol in the snow.*

Two additional proper name related errors occurred when *Grandpa* was not identified as a masculine title by engcg. Although this particular example could be easily corrected by providing an external list of titles similar to the lists of proper names, it points out how any program relying on real-world knowledge can be misled by missing information. The problems associated with proper name recognition were addressed in section 2.4 above.

Four errors occurred because AM contains no logic for processing conjunctions and therefore does not generate the correct number constraints when dealing with input such as

Susie and Molly at the playground.

They are having fun.

They is a plural pronoun, but *Susie* and *Molly* individually are singular and so AM is unable to satisfy the number constraint. Determining the proper scoping of conjunctive (and disjunctive) expressions is more properly a parsing problem than an anaphora resolution issue and was therefore not addressed here.

The remaining four errors occurred because of an animacy constraint mismatch in the context of

Close-up of a puppy chewing on a stick.

It is lying on green grass.

There are no syntactic or lexical cues here to determine that *it* has a referent which is animate (*puppy*). The three nouns in the first sentence, *close-up, puppy* and *stick* are thus left on equal footing with no meaningful way to choose between them. AM then matches by syntactic role (subject with subject) and incorrectly assigns *close-up* to *it*.

This is an extremely difficult problem to solve. It would be possible to eliminate *close-up* as a possible referent by creating a list of exceptions, along with related terms like *a picture of, a shot of,* etc. but then one could argue that this example is in fact ambiguous, since *it* could just as well refer to *stick* as *puppy*. Even performing a semantic analysis would not help in this case, since logically both a puppy and a stick could be lying on grass.

If the multiple errors with identical causes are grouped and treated as single error sources, and the errors due to engcg are excluded, the success rate becomes 10 correct and 4 incorrect, or a success rate of 71% based on type of input.

Comparing this performance to that of other anaphora resolution systems is difficult because there exists no common publicly available standard database comparable to those available for the evaluation of various image processing tasks (such as the FERET database for face detection). In fact, the design of a suitable database for such evaluation is a complex problem in its own right which is currently being addressed [Gris]. The best available programs for anaphora resolution today are claimed to have accuracy rates of 70% to 85% [Kenn], although without knowing how those numbers were obtained, it is rather difficult to make any meaningful comparisons. To the extent that such as comparison can be made, the results obtained here are not out of line. At least some further performance improvement to AM could be made easily, with a number of difficult problems remaining as open questions. Reaching 100% accuracy is probably impossible, given that some anaphoric references will always be ambiguous even to human readers. At least one such example was present even in the limited test data used herre. It should also be remembered that since a typical sentence has more than two possible referents for a given anaphor, that an accuracy of 50% for example is actually better than random choice.

A much more extensive database of 118,794 image captions from the Strong Museum in Rochester, NY (courtesy of Judy Banker) containing many pronominal anaphora is also available. The following table gives the number present of each type:

he	1012
she	943
it	1745
its	726
it's	62
him	263
his	1584
her	1636
hers	9
they	404
them	1330
their	566
they're	1

Table 3: Pronoun distribution in Strong data

Unfortunately, there was no ground truth data available for this dataset at the time of this writing and the data also needs reformatting due to system conversion so it was not used in testing.

4 Discussion

4.1 Parsing problems

AM makes extensive use of two external programs (*engcg* and *WordNet*) to gain information about the words in its input sentences. Unfortunately, AM can be no more accurate than the information it receives. Problems which arise from interacting with both engcg and WordNet are described in the following sections.

4.1.1 Problems with engcg

Some parses contain ambiguities. For example, the input

Joe took this picture of our camp house in Vermont.

returns

```
Joe took this picture of our camp house in Vermont.
"<*joe>"
      "joe" <*> <Proper> N NOM SG  @SUBJ
"<took>"
      "take" <as/SVOC/A> <for/SVOC/A> <SVO> <SVOO> <SV> V PAST VFIN
@+FMAINV
"<this>"
      "this" DET CENTRAL DEM SG @DN>
"<picture>"
      "picture" N NOM SG  @OBJ
"<of>"
      "of" PREP  @<NOM-OF
"<our>"
      "we" PRON PERS GEN PL1 @GN>
"<camp>"
      "camp" N NOM SG  @SUBJ @NN> @<P
"<house>"
      "house" N NOM SG  @<P
      "house" <SVO> <SV> V PRES -SG3 VFIN @+FMAINV
"<in>"
      "in" PREP  @<NOM @ADVL
"<*vermont>"
      "vermont" <*> <Proper> N NOM SG  @<P
"<$.>"
```

Note that it tags *house* as either a noun or a verb. Sometimes engcg is simply wrong. For example, consider the following input similar to the previous sentence:

```
Diana in the wood storage shed of our camp house in Vermont.
"<*diana>"
      "diana" <*> <Proper> N NOM SG  @SUBJ
"<in>"
      "in" PREP  @<NOM
"<the>"
      "the" <Def> DET CENTRAL ART SG/PL @DN>
"<wood>"
      "wood" N NOM SG  @NN> @<P
```

```
"<storage>"
      "storage" <-Indef> N NOM SG  @NN>
"<shed>"
      "shed" N NOM SG  @<P
"<of>"
      "of" PREP  @<NOM-OF
"<our>"
      "we" PRON PERS GEN PL1 @GN>
"<camp>"
      "camp" N NOM SG  @<P
"<house>"
      "house" <SVO> <SV> V PRES -SG3 VFIN @+FMAINV
"<in>"
      "in" PREP  @ADVL
"<*vermont>"
      "vermont" <*> <Proper> N NOM SG  @<P
"<$.>"
```

Note that here it tags *house* exclusively as a verb. Although the word *house* in fact can be used as a verb, that is not possible in the given context due to the number agreement requirement. Adding an *is* to this fragment to make it a complete sentence (*Diana* ***is*** *in the wood storage shed of our camp house in Vermont.*) does yield the noun tag for *house*, but it still has the verb tag too, creating an ambiguous parse as in the previous example.

Engcg also sometimes performs incorrect contraction expansion. For example, in

It's been very cold this spring.

the root form of the verb in *it's* is identified as *be* instead of *have*.

Engcg has particular trouble dealing with sentence fragments. For example, incorrect syntactic tags are assigned in the next example even though this fragment is missing only a leading quantifier to be a grammatically complete sentence:

```
Wood jointed elephant has glass eyes.
"<*wood>"
      "wood" <*> N NOM SG  @SUBJ
"<jointed>"
      "joint" <SVO> V PAST VFIN @+FMAINV
"<elephant>"
      "elephant" N NOM SG  @OBJ
"<has>"
      "have" <SVO> <SVOC/A> V PRES SG3 VFIN  @+FMAINV
"<glass>"
      "glass" N NOM SG  @OBJ @NN>
"<eyes>"
      "eye" N NOM PL  @OBJ
"<$.>"
```

In this case, *wood* has been assigned the subject, whereas it is really only a modifier of the true subject, *elephant*. Given the noun phrase

Friends and relatives enjoying lunch.

lunch is parsed as

```
"<lunch>"
      "lunch" <SV> V PRES -SG3 VFIN @+FMAINV
      "lunch" <SV> V IMP VFIN @+FMAINV
```

Here engcg is unsure and offers two choices. Unfortunately, both interpretations are wrong since *lunch* is actually a noun (the usage of *lunch* as a verb is comparatively uncommon). The IMP (*imperative*) tag is impossible under any interpretation. However, if this example is rephrased as a complete sentence,

Here are my friends and relatives enjoying lunch.

we now get the correct

```
"<lunch>"
      "lunch" N NOM SG  @OBJ
```

4.1.2 Problems with WordNet

AM uses the *hypernym* (part-of) relationships in the WordNet database to provide it with real-world information used to establish constraints on the nouns being considered as possible antecedents during the anaphora resolution process. For example, the second hypernym level for *girl* includes the word *female*. This can be used to establish a gender constraint.

One problem with this occurs in conjunction with nouns which have multiple senses, a situation that is extremely common. Since AM is not a complete discourse-understanding system, it has no way of telling which sense is the correct one. Therefore it examines all senses. This leads to the situation where the sentence

Patty and Sam at the birthday party.

returns *living* as a hypernym of *party*. This happens because one of the senses of *party* is ”a person” (as in “party of the first part”) and one of the attributes of persons *is living being*. Unfortunately, there is no easy way to determine which sense of a word should be used to search for constraints. For example, the first sense of *party* is *political party*, while *party* as a social occasion (which one might expect to encounter more frequently in snapshot captions) is only the second sense listed. However, the first entry for the word *kid* is the expected *child* sense, while the second sense refers to *kidskin*.

Another problem with hypernyms is that some features can be found with unintended meanings. For example, *camp* also has the feature *living*, which initially seems quite odd. In this case it is simply because the second-level hypernym for the first sense of *camp* is *living quarters*. Therefore the word *living* cannot be used to determine a constraint for animacy. However, the word *creature*, which might not be initially thought of, does seem to satisfy that requirement.

4.1.3 Abbreviations

The recognition of abbreviations poses special parsing problems because of the semantic ambiguity in the use of the period character as both an abbreviation indicator and an end-of-sentence (EOS) marker. AM uses the period as a marker to determine where one sentence in its input ends and the next begins. If it encounters an abbreviation in the first sentence, the scan will end prematurely.

One way to deal with this problem would be to maintain a dictionary of abbreviations and check the last word in the sentence being recognized against it. If a match is found, it would be assumed to be an abbreviation and scanning for the true EOS would resume. In fact, engcg does identify abbreviations so one solution would be to let engcg scan the entire input for abbreviations before performing the sentence partitioning. However, this leads to the complementary problem of incorrect abbreviation identification. For example, given a sentence ending in *"…with Ted and Reg.",* engcg incorrectly gives *Reg* the <ABBR> tag rather than the <PROPER> tag and omits the expected <$.> EOS marker entirely. The moral seems to be "you can't win".

In any event, it is unclear to what extent abbreviations may be found in real-world image captions and queries. They occur only twice in the 72 caption set used to test the current program, but they are abundant in a much larger set obtained from the Strong Museum. In any case, this problem is beyond the scope of the problem of anaphora resolution *per se* and cannot be addressed here. The practical effect on AM is that the input will be partitioned into extra "sentences" when abbreviations are encountered. This does not impede the functioning of the constraint matching process itself, although it does present problems for engcg, since that program is considerably less reliable with both part-of-speech and syntactic classification when working on sentence fragments rather than whole sentences as shown in the previous section.

4.2 Internationalization

With the global potential of the market for natural language image query and captioning, consideration should be given to the extension of automatic anaphora resolution to other languages besides English. Anaphora are a linguistic universal, occurring in all natural languages and the semantic problems associated with them are the same. The simplest way of addressing this problem would be to first translate the input text automatically into English and then proceed as usual. This approach is attractive since it requires no modification to the existing code. Unfortunately, its success is directly proportional to the accuracy of the translated text. Given the long-standing difficulties associated with machine translation, it appears that this path is unlikely to be optimal, at least for some time into the future. Constraint-matching depends heavily on having available the proper part-of-speech tags and word senses, and the correct identification of these features is among the most problematic aspects of current machine translation programs.

One alternative is to modify the anaphora resolution program. AM is written in a modular fashion so that the constraint-matching section is independent of the constraint-generation process. Unfortunately, the syntactic and morphological patterns that may be used to generate constraints for anaphora resolution can differ considerably from language to language. Some of the problems that may be encountered include:

- *Gender agreement:* All common nouns in French (as well as many other languages) have grammatical gender and require the appropriate corresponding pronouns although this gender is completely arbitrary and therefore cannot be used reliably to extract semantic information. For example *skirt* is feminine (*la jupe*) but *petticoat* (*le jupon*) is masculine. German and Russian have three grammatical genders; Swahili has at least six [Lyon]. Additional genders help improve the probabilities of arriving at a match since fewer nouns fall into each category. German has an additional complication in that all nouns are capitalized, which complicates the process of proper noun identification, and therefore gender assignment. However, perhaps as some measure of compensation, in some countries such as Germany, first names are assigned by law as being exclusively either male or female.

- *Number agreement:* Some languages have pronominal number ambiguity For example, in German *sie* can be either *she* or *they*. In fact a number agreement problem even exists within English. Some words (such as *family, group, team, corporation*) can be used as plural in British English, however, they are always singular in American English.

Of course one difficulty with this approach is the need to have a parser and lexical database equivalent to engcg and WordNet respectively available for each language for which processing is desired. This appears to be a classic case of the TANSTAAFL Principle (There Ain't No Such Thing As A Free Lunch), however, it is not an insurmountable problem. In fact, the EuroWordNet project [Voss], is currently underway and addresses Dutch, Italian, and Spanish. Parsers are also now available for many languages worldwide including French, German [Weis], Portuguese [Vill], and Chinese [Zhou].

4.3 Factors affecting accuracy

There are a number of factors beyond the control of AM which will adversely affect the accuracy of the results. Among these are the accuracy of the parse returned by engcg. As noted in the previous section, this can be quite variable. For example, it correctly identified *bldg* as an abbreviation and a noun even though the period was missing. However, it misidentified the proper name *Reg* as an abbreviation when it was the final word in a sentence and also sometimes identifies nouns as verbs. If a noun is misclassified as a verb, there is no chance it will be identified as the co-referent of any anaphora.

Spelling and typographical errors present similar problems to both engcg and WordNet. One possible solution is to run the input through a spell-checker before further processing. However, this leads to all of the problems associated with automatic spelling correction, including correction of misspelled words to the wrong word, flagging of unknown words as misspellings, inability to correct typos involving incorrect placement of space characters (e.g. *whic hword*), inability to recognize misspellings that result in valid words, but the wrong word such as *food* vs. *good* (*f* is adjacent to *g* on the keyboard), and inability to correct homophones (many people do not seem to know the difference between *site, sight,* and *cite* or *there, their,* and *they're*.). Spell-checking also cannot address such problems as typos where a sentence is mistakenly ended with a comma rather than a period (the two characters are adjacent on the keyboard) and unconventional capitalization. These problems all occur in practice and need to be anticipated in any application destined for use by a general audience, as opposed to professional captioners trained to operate in a specific style. All of these problems will also occur and even more frequently if voice captioning should be implemented since automatic speech-to-text programs have many of the same problems as machine translation programs.

Another problem with both WordNet and engcg is that of dealing with unknown words. This can arise either because of misspellings as discussed above, or because the word is truly unknown. If one or both of the tagger and lexicon fail to identify the part of speech of a word, it could lead to a matching failure if that word turns out to be a noun that is an anphoric referent. Although none of the words in the test database were unknown, it is not difficult to think of words that neither engcg nor WordNet know. At least engcg will take a guess as to the part of speech of an unknown word based on its syntactic placement. For example, in

The fnord is red.

it thinks *fnord* is a noun and marks it with <?> to indicate that this is only a guess.

However, WordNet has no way of dealing with unknown words. This is a difficult problem that is beyond the scope of AM, but proposals have been made for how to handle it [Mikh] and this should be kept in mind as needing attention to increase the accuracy of results.

5 Summary

Automatic resolution of anaphora by computer is a difficult natural language processing problem that requires considerable knowledge beyond the information which is available in the input text to solve accurately. Despite considerable effort on the part of researchers in computational linguistics, it remains an open problem. The best programs at this time achieve no more than 75% to 85% accuracy. However, anaphora resolution is worth considering because it offers the possibility of increasing the accuracy of matches in a natural language image database query system which could provide a competitive advantage to the operator of such a system by increasing customer satisfaction. As the size of such databases increases in the future, the use of systems which only match keywords or attempt to perform semantic expansion (searching for terms semantically related to those given in the query) will become less acceptable.

The program presented here is capable of performing intersentential pronominal anaphora resolution on ordinary English text and includes a comprehensive treatment of pleonastic pronouns, something which almost all other investigators have ignored, but which I believe is crucial for successful treatment of anaphora.

Many opportunities remain for extending the current work. Different taggers of English could be tried, e.g. TTP [Strz] or TOSCA/LOB.. A more extensive test database with ground truth analysis should be implemented. Finally, the various unresolved issues raised in the discussion could be addressed. At this date, these remain the subject of active research in the natural language processing community. While perfection is probably an unreasonable goal, it would seem that a considerable improvement in resolution accuracy could be obtained for a relatively small additional investment in time simply by adding the unimplemented features described here to the existing code.

Finally, given the recent growing interest in natural language processing and its commercial implications, and the wide applicability of anaphora resolution in natural language processing, it is recommended that consideration be given to obtaining a patent position in this area.

6 Appendix

All file names and URL's included in this section are known to be current at the time of this writing. All programs and data are in the ISL Unix environment.

6.1 Code and data location

AM is implemented entirely in Interlisp as contained in Medley Lisp, the successor to Interlisp-D. The source code is found in the file `/space/cloves1/lisp/lyric/VIMFNS`. The top-level function is called `AM` and it takes one argument: a string which may be either a set of sentences containing the anaphora to be resolved or a file name which contains these sentences. AM internally calls engcg and WordNet.

engcg is found in `/home/bhandari/engcg/engcg-971001/engcg` and can be run from the Unix command line. It takes one argument: a file name containing the input sentences to be parsed. It will only run on the server *natasha* because of the program's licensing arrangement. A wrapper script called `eng.perl` is available in `/space/cloves1/perl/` which allows input to be typed in directly from the command line and can be run on any machine in the ISD computing environment. To use it, alias `eng` to `rsh natasha /space/cloves1/perl/eng.perl`. Then type `eng <CR>` followed by the desired text on the next line. The freeware TOSCA/LOB tagger is available on `/space/cloves1/vim/tlbtag`, however, this only runs in MS-DOS.

WordNet is installed in `/freeware/misc/bin/wordnet/wordnet-1.6/bin/`. It will run on any machine and can either be called from the command line with the command `wn` or via a graphical user interface with the command `wnb`.

The world knowledge proper name data files used by AM are located in `/home/bhandari/nl/mit/nlp/` (`male-names.data`, `female-names.data` and `place-names.data`). This directory also contains a small list of abbreviations, however, these are currently not used by AM. The intersections of the name files are in `/space/raccoon2/vim/data/`.

The caption data files used for development and testing are located in `/space/project3/gray/imagecaptions/`. This directory contains data received from the Strong Museum, the George Eastman House, as well as locally developed captions based on the JBJL image database. The JBJL images are located in `/space/library/jbjl/`.

6.2 WWW resources

Lingsoft, the company that produces engcg has a web site with documentation on their product at http://www.lingsoft.fi. There is no locally available documentation on engcg.

The home page for WordNet is at http://www.cogsci.princeton.edu/~wn/. This site contains documentation and also lets you use WordNet interactively. There is a corresponding site for EuroWordNet at http://www.let.uva.nl/~ewn.

PNI (Picture Network International) operates the Publisher's Depot web site which supports natural language queries for locating stock photography, although without anaphoric resolution. It may be viewed at http://www.publishersdepot.com. Another search engine using natural language input for queries is MetaSEEk at http://www.ctr.columbia.edu/metaseek.

The reader interested in pursuing the references further may be interested in http://gopher.sil.org/reference which is an extensive glossary of technical linguistic jargon. The web site of the Association for Computational Linguistics is http://www.aclweb.org.

7 Patents

Interestingly, a search of the U.S. Patent Database (searching all fields, back to 1971) revealed *zero* patents mentioning the term *anaphora*, and only five that used the term *pronoun*. Of those five, only one could be considered remotely relevant to anaphora resolution [Tana].

This is a natural language understanding patent that implements a reverse question answering scenario (the computer asks the questions, the human user provides answers). At first glance, it appears that they are doing anaphora resolution, based on the following sentence in the abstract:

> "The inputting and editing system analyzes the answer to an interrogative and matches the pronouns and other elements of the answer with their corresponding elements from the interrogative so that the operator does not have to answer the interrogative with the exact language of the interrogative for the system to understand and utilize the answer properly."

However, examination of the actual claim pertaining to this section reveals the following:

> "using said inferences of step (e) [*developing inferences based on the contents of the sentence*] to obtain a case nominal corresponding to a pronoun used in the interrogative;"

So this does not really appear to be anaphora resolution at all. They have the computer generate a query which already contains anaphora, so (presumably) the computer already knows to whom or what the pronouns it is outputting refers to. Therefore they simply have to look for a noun in the user's response that has characteristics similar to the noun it already knows about (and replaced with a pronoun). This is pattern matching, not anaphora resolution. In any event, their claim is strictly limited to question answering.

There is one other patent is worth noting [Kaji]. This is a machine translation system which apparently performs ellipsis identification in the source language and then inserts anaphora into the translation, as stated in their abstract:

> "When a target language translation of the source language sentence is formed, a pronoun having the same gender, person and number as the omitted phrase is used as a target language equivalent for the omitted phrase, and thus a target language translation which is grammatically correct, is obtained."

This of course is also not anaphora resolution.

There actually are many patents on the general topic of systems for retrieving images using natural language. A recent search using the terms *natural language image retrieval library* returned 1000 patents, although only perhaps a dozen are truly relevant to image query by natural language. Many of these patents deal with limited applications, such as question-answering systems for specific domains, and address the natural language component only tangentially providing very little detail as to its actual operation. Perhaps the most relevant patent in this group is one assigned to Matshushita in 1997

[Okam]. It appears to address the major aspects of an image query system including parsing of the input data, semantic analysis using Fillmore's case grammar [Fill], constraint matching, conversion of caption data into an intermediate form, and generating confidence scores for the results. It goes so far as to present a simple grammar of English used to analyze input sentences. However, it specifically does not appear to address anaphora in either the detailed description or the claims, and its rules for noun phrase expansion do not include pronouns.

8 Bibliography

The bibliography is divided into two sections. The first contains all citations in the text not directly related to anaphora resolution including US patents. The second is a general survery of the open literature on anaphora resolution.

8.1 Citations

Allen, James, *Natural Language Understanding*, Benjamin/Cummings, Redwood City, Ca., 1995

Bhandari, Archna, *A Method for using Natural Language for the Description, Search, and Retrieval of Multi-media Objects*, Eastman Kodak Internal Technical Report, Aug. 22, 1997

Carroll, Lewis, *Alice's Adventures in Wonderland*, in *The Annotated Alice*, World Publishing, New York, 1972

Chomsky, Noam, *Aspects of the Theory of Syntax*, MIT Press, Cambridge, 1965

Culicover, Peter W., Ray Jackendof, "Something else for the binding theory", *Linguistic Inquiry*, **26**:2, 249-275, Spring 1995

Fellbaum, Christiane, ed., *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, 1998

Fillmore, Charles, "The case for case", E. Bach, R.T. Harms, eds.,*Universals of Linguistic Theory*, Holt, New York, 1968

Garey, Michael R., David S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W.H. Freeman, San Francisco, 1979

Grishman, Ralph, Beth Sundheim, "Message Understanding Conference - 6: A Brief History", *Proc. 16th International Conference on Computational Linguistics*, Copenhagen, Denmark, August 1996

Grishman, Ralph, Beth Sundheim, ""Design of the MUC-6 Evaluation", *Advances in Text Processing: Tipster Program Phase II*, Morgan Kaufmann 1996

Haase, Kenneth, "FramerD: Representing knowledge in the large", *IBM Systems Journal*, **35**:3&4, 381-397, 1996

Kaji, Hiroyuki, Y. Nitta, *System For Automatic Language Translation Using Several Dictionary Storage Areas And A Noun Table,* U.S. Patent no. 5,020,021, assigned to Hitachi, Ltd., Tokyo, Japan, issued May 28, 1991

Karlsson, F., A. Voutilainen, J. Heikkilä, A. Anttila, *Constraint Grammar. A Language-Independent System for Parsing Unrestricted Text*, Mouton de Gruyter: Berlin and New York, 1995

Lyons, John, *Introduction to Theoretical Linguistics*, Cambridge University Press, 1971
Pages 405-423

Mikheev, Andrei, "Automatic rule induction for unknown-word guessing", *Computational Linguistics*, **23**:3, 405-423, June 1997

Molière, *Le Bourgeois Gentilhomme*, in *The Miser and Other Plays*, Penguin Books, 1987

Okamoto, Shusaku, *Apparatus And A Method For Retrieving Image Objects Based On Correlation With Natural Language Sentence Parameters*, US Patent no. 5,684,999, assigned to Matsushita Electric Industrial Co., Ltd., Kadoma, Japan, issued Nov. 4, 1997

Pirkola,A. & Järvelin,K, "Recall and precision effects of anaphor and ellipsis resolution in proximity searching in a text database", in: Ingwersen, P. & N.O. Pors, ed., *Proceedings CoLIS, 2nd International Conference on Conceptions of Library and Information Science: Integration in Perspective*, Copenhagen, Oct. 13-16, 1996, Copenhagen: The Royal School of Librarianship, 459 – 475, 1996

Strzalkowski, Tomek, "TTP: A fast and robust parser for natural language", *Proc. 14th Int'l Conf. on Computational Linguistics* (COLING 92), Nantes, France, July 23-28, 1992

Tanaka, Toshiyuki , S. Kuga, *et al., Inputting and Editing System in a Knowledge Based Inquiry and Answer System*, US Patent no. 4,837,689, assigned to Sharp Kabushiki Kaisha, Osaka, Japan , issued Jun. 6, 1989

Villavicêncio, Aline, Nuno Marques, J.G.P. Lopes, Fábio Villavicêncio, "Part-of-speech tagging for Portuguese texts", *Proc. 12th Brazilian Symposium on Artificial Intelligence*, Campinas, Springer Verlag, October 11-13, 1995

Vossen, Piek, "EuroWordNet: a multilingual database for information retrieval", *DELOS Workshop on Cross-language Information Retrieval*, Zurich, March 5-7, 1997

Weisweber, Wilhelm, *Ein Dominanz-Chart-Parser für generalisierte Phrasenstrukturgrammatiken*, KIT-Report 45, Institute for Software and Theoretical CS, Technical University of Berlin, 1987

Zhou, Qiang, Changning Huang, "A Chinese syntactic parser based on bracket matching principle", *Comm. Chinese and Oriental Languages Information Processing Society* (COLIPS), **7**:2, 53-60, December 1997

8.2 Literature Survey

Most of the following references are based on a number of online bibliographies including one for a tutorial on anaphora resolution recently given at *COLING-ACL '98*. Although most are not cited in the text they are included here for completeness. They form a reasonably comprehensive view of the treatment of anaphora resolution in the open literature over the last 25 years. For an evaluation of the applicable patent literature, see section 7.

A

Abraços, José, "Pronominal anaphora resolution", Master's Thesis, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 1993

Abraços, José, J.G.~Pereira Lopes, "Extending DRT with a focussing mechanism for pronominal anaphora and ellipsis resolution", *Proc. COLING '94*, 1128-1132, 1994

Aone, C., D. McKee, 1993, "A language-indepentend anaphora resolution system for understanding mutlilingual texts", *Proc. 31st Annual Meeting of the Association for Computational Linguistics*, 156-163, Columbus, Ohio, 1993

Aoun, J., "A symmetric theory of anaphoric relations", P. Sells & C. Jones, eds., *Proc. North East Linguistic Society (NELS) 13*, Université du Québec à Montréal, GLSA Publications, 1983

Asher, Nicholas, Hajime Wada, "A computational account of syntactic, semantic and discourse principles for anaphora resolution", *J. Semantics*, **6**, 309-344, 1988

Asher, Nicholas, "Reference to abstract objects in discourse", *Studies in Linguistics and Philosophy*, **50**, Kluwer, Dordrecht, 1993

Austin, Peter, Joan Bresnan, "Non-configurationality in Australian aboriginal languages", *Natural Language and Linguistic Theory,* **14**:2, 215-268, May 1996

Avrutin, Sergey, Maria Babyonyshev, "Obviation in subjunctive clauses and AGR: Evidence from Russian", *Natural Language and Linguistic Theory,* **15**:2, 229-262, May 1997

Azzam, Saliha, "Anaphors, PPs and disambiguation process for conceptual analysis", *Proc. 14th IJCAI*, 1995

Azzam, Saliha, "An algorithm to co-ordinate anaphora resolution and PPS disambiguation process", *Proc. 7th Conference of the European Chapter of the Association for Computational Linguistics* (EACL), University College Dublin, Ireland, March 27-31, 1995

Azzam, Saliha, "Resolving anaphors in embedded sentences", *Proc. 34th Annual Conference of the ACL*, 1996

Azzam, Saliha, Kevin Humphreys, Robert Gaizauskas, "Coreference resolution in a multiligual information extraction system", *Proc. 1st Language Resources and Evaluation Conference* (LREC), Linguistic Coreference Workshop, 1998

Azzam, Saliha, Kevin Humphreys, Robert Gaizauskas, "Evaluating a focus-based approach to anaphora resolution", *Proc. COLING-ACL '98*, Montreal, Canada, 74-78, Aug. 9-14, 1998

B

Baldwin, B., "Anaphora resolution with centering", M. Walker, E. Prince, A. Joshi, eds., *Centering Theory in Discourse*, Oxford University Press, 1997

Baldwin, B., *CogNIAC: a discourse processing engine*, Ph.D. dissertation, University of Pennsylvania, Department of Computer and Information Sciences.

Baldwin, B., "CogNIAC: high precision coreference with limited knowledge and linguistic resources", *Proc. ACL '97/EACL '97 Workshop on Operational Factors in Practical, Robust Anaphora Resolution*, Madrid, July 11, 1997

Balthazart, D., L. Kister, "Is it possible to predetermine a referent included in a French *N de N* structure?", *Approaches to Discourse Anaphora: Proceedings of the Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Barros, Fl., A. Deroeck, "Resolving anaphora in a portable natural language front end to databases", *Proc. 4th Conference On Applied Natural Language Processing*, Stuttgart, Germany, October 13-15, 1994

Barwise, J., "Noun phrases, generalized quantifiers and anaphora", P. Gardenfors, ed., *Generalized Quantifiers: Linguistic And Logical Approaches*, 1-30, D. Reidel Publishing Company, Dordrecht, Germany, 1987

Bender, A., G. Dogil, J. Mayer: Prosodic Disambiguation of Anaphoric Pronouns in German Discourses", *Approaches to Discourse Anaphora: Proceedings of the Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Bersani Berselli, Gabriele, "Nominal ambiguity and disambiguation by anaphoric devices", *Approaches to Discourse Anaphora: Proceedings of the Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Bergler, Sabine, “Towards reliable partial anaphora resolution”, *Proc. ACL '97/EACL '97 Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, Madrid, Spain, July 11, 1997

Berthonneau, Anne-Marie, Georges Kleiber, “Pour une nouvelle approche de l'imparfait: l'imparfait, un temps anaphorique méronomique”, *Langages*, December 1993

Berwick, R.C., A. Weiberg, *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition,* MIT Press, Cambridge, 1984

Bosch, Peter, “Some good reasons for shallow pronoun processing”, *Proc. IBM Conference on Natural Language Processing*, Thornwood, New York, 1988

Boguraev, B.K., *Automatic Resolution of Linguistic Ambiguities*, University of Cambridge Computer Laboratory TR-13, 1979

Bowden, Paul R., Peter Halstead, Tony G., “Rose: Endophor resolution in a pattern-matching knowledge extraction system”, *Approaches to Discourse Anaphora: Proceedings of the Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Branco, Antonio, “Irreflexive zooming: reciprocals processing as indirect anaphor resolution”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Brennan, Susan E., Marilyn W. Fridman, Carl J. Pollard, “A centering approach to pronouns”, *Proc. 25th Annual Meeting of the Association for Computational Linguistics (ACL*), 155-162, Stanford, CA, 1987

Broadwell, G. A., “A-bar anaphora and relative clauses”, S. Berman, J.W. Choe, J. McDonough, eds., *Proc. North East Linguistic Society (NELS) 16*, McGill University, GLSA Publications, 1986

Bullwinkle, Candace, “Levels of complexity in discourse for anaphora disambiguation and speech act interpretation”, *Proc. 5th IJCAI*, Cambridge, Massachusetts, 1977

C

Cahn, Janet, “The effect of intonation on pronoun referent resolution”, *Proc. ACL ‘95*, 290-292, 1995

Carbonell, Jaime G., Ralf D. Brown, “Anaphora resolution: a multi-strategy approach”, *Proc. 12th International Conference on Computational Linguistics COLING '88*, 96-101, Budapest, August, 1988

Carter, D.M., *A shallow processing approach to anaphor resolution*, Technical Report No. 88 (Ph.D. thesis). University of Cambridge, Computer Laboratory, 1985

Carter, D.M., "Common-Sense Inference in a Focus-Guided Anaphor Resolver", *J. Semantics*, **4**, 237-246, 1987

Carter, D.M., *Interpreting Anaphora in Natural Language Texts*, Ellis Horwood, Chichester, England, 1987

Carvalho, A., "Logic Grammars and pronominal anaphora", *Approaches to Discourse Anaphora: Proceedings of the Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Chanier, Thierry, "Une architecture collaborative pour la résolution d'anaphores dans des textes techniques écrits en langage naturel", *Revue d'intelligence artificielle*, **3**:3, 41-65, 1989

Chen, Hsin Hsi, "The transfer of anaphors in translation", *Literary and Linguistic Computing*, **7**:4, 1992

Chen, Hsin Hsi, "Anaphora resolution algorithms for Mandarin Chinese", *Comm. Chinese and Oriental Languages Information Processing Society* (COLIPS), **3**:2, December 1993

Cheng, Lisa L.-S., C.-T. James Huang, "Two types of donkey sentences", *Natural Language Semantics,* **4**:2, 121-163, 1996

Chierchia, G., "Anaphora and dynamic binding", *Linguistics and Philosophy*, **15**:2, 111-183, 1991

Chung, Chan, "Scrambling in Korean and its effect on anaphor binding: an alternative to movement approaches", *Proc. ESCOL '93* (Eastern States Conference on Linguistics), CLC Publications, Cornell University, Ithaca, NY, 1994

Chur, J., "Generic anaphora in German texts", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Clark, Herbert H., C.J. Sengul, "In search of referents for nouns and pronouns", *Memory and Cognition*, **7**, 35-41, 1979

Cohen, Phillip R., "The pragmatics of referring and the modality of Communication", *Computational Linguistics*, **10**, 97-146, 1984

Coelho, O.B., “A quasi-connectionist approach to anaphora resolution in task-oriented discourses”, *Proc. AAAI ‘92 Workshop on Integrating Neural and Symbolic Processes - the Cognitive Dimension*, San Jose, CA, July 12-16, 1992

Coelho, O.B., “A connectionist approach to anaphora resolution in task-oriented discourses”, I. Aleksander, J. Taylor, eds., *Artificial Neural Networks II: Proc. 2nd International Conference on Artificial Neural Networks*, Elsevier, Brighton, England, September 1992

Connoly, D., J. Burger, D. Day, “A machine learning approach to anaphoric reference”, *Proc. International Conference on New Methods in Language Processing,* UMIST, Manchester, September 14-16, 1994

Conte, Maria-Elisabeth, “Facts, events, propositions in anaphoric encapsulation”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Cooper, Robin, “The interpretation of pronouns”, Heny, Schnelle, eds., *Selections From the 3rd Gronigen Round Table, Syntax and Semantics,* **10**, Academic Press, NewYork, 1979

Corblin, Francis, “L’anaphore en subordination modale”, *Actes du colloque d’Anvers: Relations anaphoriques et (in)cohérence,* 1994

Cosse, Michel, “Indefinite associative anaphora in French”, *Approaches to Discourse Anaphora: Proceedings of the Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Crain, S., M.J. Steedman, “On not being led up the garden path: the use of context by the psychological syntax processor”, Dowty, Karttunen, Zwicky, eds., *Natural Language Parsing*, Cambridge University Press, Cambridge, 320-358, 1985

Crain, S., C. McKee, “Acquisition of structural restrictions on anaphora”, S. Berman, J.W. Choe, J. McDonough, eds., *Proc. North East Linguistic Society (NELS) 16*, McGill University, GLSA Publications, 1986

Crawley, R.A., R. Stevenson, “Reference in single sentences and in texts”, *J. Psycholinguistic Research*, **19**, 191-210, 1990

Crawley, R.A., R. Stevenson, D. Kleinman, “The use of heuristic strategies in the comprehension of pronouns”, *J. Psycholinguistic Research*, **19**, 245-264, 1990

D

Dagan, Ido, A. Itai, “Automatic processing of large corpora for the resolution of anaphora references”, *Proc. 13th International Conference on Computational Linguistics*, COLING '90, Helsinki, 1990

Dagan, Ido, John Justeson, Shalom Lappin, Herbert Leass, Amnon Ribak, “Syntax and lexical statistics in anaphora resolution”, *Applied Artificial Intelligence*, 1994

Dahl, D., C. Ball, *Reference resolution in PUNDIT,* Research Report CAIT-SLS-9004,. Center for Advanced Information Technology, Paoli, PA 19301, March 1990

Dalrymple, Mary, “The syntax of anaphoric binding”, *CSLI Lecture Notes no. 36*, Center for the Study of Language and Information, Stanford, Ca.,1993

Dalrymple, Mary, John Lamping, Fernando Pereira, Vijay Saraswat, “Quantifiers, anaphora, and intensionality”, *J. Logic, Language and Information*, **6**:3, 219-273, July 1997

Danielsson, Pernilla, *Anaphora Resolution in Swedish Texts for Machine Translation*, Master's Thesis, Dept. of Computational Linguistics, University of Gothenburg, Sweden, 1995

DeChicchis, Joseph , “Mayan verb complex anaphors”, *Proc. ESCOL ’88* (Eastern States Conference on Linguistics), Ohio State University, Columbus, Ohio, CLC Publications, 1989

Dekker, Paul, “Predicate logic with anaphora”, Mandy Harvey, Lynn Santelmann, eds., *Proc. from Semantics and Linguistic Theory IV* (SALT IV), University of Rochester, Rochester, NY, CLC Publications, May 6-8, 1994

Di Eugenio, Barbara, “Centering theory and the Italian pronominal system”, M. Walker, E. Prince, A. Joshi, eds., *Centering theory in discourse*, Oxford University Press, 1997

Dotzek, K., *Strategien zur Pronominalisierung*, IBM Wissenschaftliches Zentrum, Institut für Wissensbasierte Systeme, IWBS Report 110, Heidelberg, Germany, 1990

Dousaka, K., “Identifying the referents of Japanese zero-pronouns based on pragmatic condition interpretation”, *Tr. Information Processing Society of Japan*, **35**:5, 768-778, 1994 (in Japanese)

Downing, P., “The anaphoric use of classifiers in Japanese”, P. Downing, M. Noonan, eds., *Word Order in Discourse, Vol. 30 of Typological Studies in Language*, 345-375, John Benjamins, 1995

Dunin-Keplicz, Barbara, “Towards better understanding of anaphora”, *Proc. EACL '83*, Pisa, Italy, 139-144, 1983

Dunin-Keplicz, Barbara, “Default reasoning in anaphora resolution”, *Proc. ECAI '84*, Pisa, Italy, 157-166, 1984

Dunin-Keplicz, Barbara, “How to restrict ambiguity of discourse, *Proc. EACL '85*, Geneva, Switzerland, 93-97, 1985

Dunin-Keplicz, Barbara, “Partial reconstruction of coreferential structure of discourse”, *Proc. ECAI '88*, Munich, Germany, 732-737, 1988

Dunin-Keplicz, Barbara, *A Formal Method of Determining Coreferentiality of Pronouns in Polish*, Ph.D. Thesis (in Polish), Jagiellonian University, Crakow, Poland, 1990

Dunin-Keplicz, Barbara, “A formal treatment of referential relations”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Durrant-Peatfield, M., W. Marslen-Wilson, “Pragmatic effects on zero-anaphor assignment”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

E

Eberle, Kurt, Walter Kasper, “Tenses as anaphora”, *Proc. 4th Annual Conference of the European Chapter of the ACL*, 43-50, 1989

Elworthy, David A. H., “A theory of anaphoric information”, *Linguistics and Philosophy,* **18**:3, 297-332, June 1995

Evans, Frederic, “Binding into anaphoric verb phrases”, *Proc. ESCOL ’88* (Eastern States Conference on Linguistics), Ohio State University, Columbus, Ohio, CLC Publications, 1989

F

Fischer, I., B. Geistert, G. Görz, “Chart-based incremental semantics construction with anaphora resolution using lambda-DRT”, *Proc. 4th International Workshop on Parsing Technologies*, 87-88. Sept. 20-24, 1995

Fischer, I., B. Geistert, G. Görz, “Incremental anaphora resolution in a chart-based semantics construction framework using λ-DRT”, S. Botley, J. Glass, eds., *Proc. Discourse Anaphora and Anaphora Resolution Colloquium (DAARC ’96)*, 235-244, Lancaster, UK, July 1996

Fox, Barbara A., *Discourse Structure and Anaphora: Written and Conversational English*, Cambridge University Press, Cambridge, 1987

Fraurud, Kari, “Indirect anaphora in a cross-linguistic perspective”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Frederking, R., M. Gehrke - *Resolving anaphoric references in a DRT-based dialogue system: Part 2: Focus and Taxonomic inference*. Siemens AG, WISBER, Bericht Nr.17, 1987

Freitas, Sérgio, "Deitic and pronominal anaphora in dialogues", Master's Thesis, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil, 1993 (in Portuguese)

Fretheim, Thorstein, "Indirect anaphors and pro-verbs of the 'happen' type", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

G

Gaiffe, Bertrand, Anne Reboul, Laurent Romary, "Les NP définis: anaphore, anaphore associative et cohérence", *Actes Colloque Relations anaphoriques et incohérence*, Université d'Anvers, 1994

Gaizauskas, R., K. Humpreys, "Quantitative evaluation of coreference algorithms in an information extraction system", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Gaizauskas, R., K. Humphreys, "Quantitative evaluation of coreference algorithms in an information extraction system", S. Botley, T. McEnery, eds., *Discourse Anaphora and Anaphor Resolution*, University College London Press, forthcoming

Garrod, S., Tony Sanford, "Referential processes in reading: Focussing on roles and individuals", D.A. Balota, G.B.F. d'Arcais & K. Rayner, eds., *Comprehension*, 465-486, Lawrence Erlbaum Associates, Hillsdale, N.J., 1990

Garrod, S., D. Freudenthal, E. Boyle, "The role of different types of anaphor in on-line resolution of sentences in a discourse", *Journal of Memory and Language*, **33**, 39-68, 1994

Garrod, S., "Resolving pronouns and other anaphoric devices: The case for diversity in discourse processing", C. Clifton, Jr., L. Frazier & K. Rayner eds., *Perspectives on Sentence Processing*, 339-359, Lawrence Erlbaum Associates, Hillsdale, N.J., 1994

Gawron, M., S. Peters, "Anaphora and quantification in situation semantics", *CSLI Lecture Notes no. 19*, CSLI, Stanford, CA, 1990

Geistert, B., *Auflösung anaphorischer Referenzen in der λ-DRT*, Bachelor's Thesis, Univ. Erlangen-Nuremberg, IMMD VIII, 1995

Gordon, Peter C., Barbara J. Grosz, Laura A. Gilliom, "Pronouns, names and the centering of attention in discourse", *Cognitive Science*, **17**:3, 311-348, 1993

Greene, S.B., G. McKoon, and R. Ratcliff, "Pronoun resolution and discourse models", *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **18**, 1992

Grosz, Barbara J., Aravind K. Joshi, Scott Weinstein, "Centering: a framework for modeling the local coherence of discourse", *Computational Linguistics*, **21**, 1995

Grosz, Barbara J., Aravind Joshi, Scott Weinstein, "Providing a unified account of definite noun phrases in discourse", *Proc. 21st Annual Meeting of the Association for Computational Linguistics,* 44-50, 1983

Guenthner, Franz, Hubert Lehmann, "Rules for pronominalization", *Proc. 1st Annual Meeting of the European Chapter of the ACL* (EACL), 144-151, 1983

Guindon, R., "Anaphora resolution: short-term memory and focusing", *Proc. 23rd Annual Meeting of the ACL*, 218-227, 1985

Gundel, Jeanette, Nancy Hedberg, Ron Zacharski, "Cognitive status and the form of referring expressions in discourse", *Language*, **69**:2, 274-307, 1993

Gundel, Jeanette, Nancy Hedberg, Ron Zacharski, "Cognitive status, concreteness and the form of indirect anaphors", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

H

Hahn, U., M. Strube, "Centering in-the-large: Computing referential discourse segments", *Proc. 35th Annual Meeting of the ACL,* 104-111, 1997

Hajikova, E., V. Kubon, P. Kubon, "Stock of shared knowledge: a tool for solving pronominal anaphora", *Proc. COLING '92*, **1**, 127-133, Nantes, France, 1992

Halpern, A., G. Mauner & M. K. Tannenhaus Priming of structural and conceptual verb phrase anaphors", A. Schafer, ed., *Proc. NELS 23* (North East Linguistic Society), University of Ottawa, Canada, GLSA Publications, 1993

Hankamer, Jorge, Ivan Sag, "Deep and surface anaphora", *Linguistic Inquiry*, **7**, 391-428, 1976

Hauenschild, Christa, "Anapherninterpretation in der Maschinellen Übersetzung", KIT-Report 94, 50-66, Institute for Software and Theoretical CS, Technical University of Berlin 1992, and *Zeitschrift für Literaturwissenschaft und Linguistik,* **84**, Vandenhoeck & Ruprecht, 1991

Hayes, P.J., "Anaphora for limited domain systems", *Proc. 7th IJCAI*, Vancouver, Canada, 1981

Hazout, Ilan, “Comparative ellipsis and logical form”, *Natural Language & Linguistic Theory,* **13**:1, 1-37, February 1995

Heim, I., ”E-Type pronouns and donkey anaphora”, *Linguistics and Philosophy*, **13**:2, 137-177, 1990

Hellan, L., *Anaphora in Norwegian and the Theory of Grammar*, Foris, Dordrecht, 1988

Hinrichs, Erhard W., “Temporal anaphora in discourses of English”, *Linguistics and Philosophy*, **9**, 63-82, 1986

Hintikka, J., J. Kulas, *Anaphora and Definite Descriptions: Two Applications of Game-Theoretical Semantics*, Reidel, Dordrecht, Germany, 1985

Hirst, G.J., *Anaphora in Natural Language Understanding*, Springer Verlag, Berlin, Germany, 1981

Hirst, G.J., “Discourse-oriented anaphora resolution in natural language understanding: a review”, *American J. Computational Linguistics*, **7**:2, 85-98, 1981

Hirst, G.J., *Semantic Interpretation & the Resolution of Ambiguity*, Cambridge University Press, 1987

Hitzeman, Janet, Chris Mellish, Jon Oberlander, “Dynamic generation of museum web pages: the intelligent labelling explorer”, *Archives and Museum Informatics*, **11**:2, 107-115, 1997

Hitzeman, Massimo Poesio, “Long distance pronominalisation and global focus”, *Proc. COLING-ACL ’98*, Montreal, Canada, 550-556,1998

Hobbs, Jerry R., *Pronoun resolution*, Research Report 76-1, Department of Computer Science, City College, City University of New York, August 1976

Hobbs, Jerry R., “Resolving pronoun references”, *Lingua*, **44**, B11-338, 1978, also in Barbara Grosz, K. Sparck-Jones, Bonnie Webber, eds., *Readings in Natural Language Processing*, 339-352, Morgan Kaufmann, Los Altos, Ca., 1986

Hobbs, Jerry R., “Coherence and co-reference”, *Cognitive Science*, **3**:1, 67-90, 1979

Huang. Y., “Anaphora in sentence and in discourse: a neo-Gricean pragmatic approach”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Hudson-D'Zmura, S., *The structure of discourse and anaphor resolution: the discourse center and the roles of nouns and pronouns*, Ph.D. dissertation, University of Rochester, Rochester, NY, 1988

Hudson-D'Zmura, S., “Assigning antecedents to ambiguous pronouns: the role of the center of attention as the default assignment”, M. Walker, E. Prince, A. Joshi, eds., *Centering Theory in Discourse*, Oxford University Press, 1997

Huls, Carla, Wim Claassen, Edwin Bos, “Automatic referent resolution of deictic and anaphoric expressions”, *Computational Linguistics*, **21**:1, 59-79, March 1995

I

Iida, Masayo, *Context and Binding in Japanese*, Ph.D. Thesis, Stanford University, Linguistics Dept., Stanford, Ca. 1992

Ingria, R., D. Stallard, “A computational mechanism for pronominal reference”, *Proc. 27th Annual Meeting of the ACL,* 262-271, Vancouver, British Columbia, June 26-29, 1989

J

Johnson, M., E. Klein, “Discourse, anaphora and parsing”, *Proc. 11th Conf. On Computational Linguistics*, 1986

K

Kameyama, Megumi, “A property-sharing constraint in centering”, *Proc. 24th Annual Meeting of the ACL*, New York, 1986

Kameyama, Megumi, “Japanese zero pronominal binding, where syntax and discourse meet”, William Poser, ed., *Papers from the 2nd International Workshop on Japanese Syntax*, 47-74, Stanford CSLI, 1988, also University of Pennsylvania TR MS-CIS-86-60

Kameyama, Megumi, “Recognizing referential links: an information extraction perspective”, *Proc. ACL '97/EACL '97 Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts,* 46-53, Madrid, Spain, July 11, 1997

Kantor, R.N., *The Management & Comprehension of Discourse Connection by Pronouns in English*, Ph.D. Thesis, Department of Linguistics, Ohio State University, Ohio, 1977

Karttunen, Lauri, “Discourse referents”, James D. McCawley, ed., *Syntax and Semantics: Notes from the Linguistic Underground*, **7**, 363-386, Academic Press, New York, 1976

Kasper, W., H. Zevaat, M. Moens, “Anaphora Resolution”, G. Bes, ed., *The Construction of a Natural Language and Graphics Interface. Results and Perspectives from the ACORD Project, ESPRIT Research Reports*, Springer, Heidelberg, Germany, 1992

Katada, F., “What can long-distance anaphora say about operator systems of syntax?”, J. Carter, R.M. Dechaine, eds., *Proc. North East Linguistic Society (NELS) 19,* Cornell University, Ithaca, NY, GLSA Publications, 1989

Kawtrakul, Asanee, "Anaphora resolution based on dynamic context model in database-oriented discourse", *J. Association of Artificial Intelligence* (Japan), **5**:6, 100-113,1989

Kawtrakul, Asanee, Yasuyoshi Inagaki, "A context model approach to anaphora resolution in database-oriented discourse", *J. Association of Artificial Intelligence* (Japan), **5**:6, 49-62, Nov.1990

Kawtrakul, Asanee, *Anaphora Resolution Based on Context Model Approach in Database-Oriented Discourse*, PhD thesis, School of Engineering, Nagoya University, 1991

Keenan, Edward L., "Anaphor-antecedent assymetry: A conceptual necessity?", *Proc. from Semantics and Linguistic Theory III* (SALT III), Cornell University, Ithaca, NY, CLC Publications, 1993

Kehler, Andrew, "A discourse copying algorithm for ellipsis and anaphora resolution", *Proc. 6th Conference of the European Chapter of the Association for Computational Linguistics* (EACL-93), 203-212, Utrecht, April, 1993.

Kehler, Andrew, "Intrasentential constraints on intersentential anaphora in centering theory", *Workshop on Centering Theory in Naturally Occurring Discourse*, University of Pennsylvania, May 1993

Kehler, Andrew, "The effect of establishing coherence in ellipsis and anaphora resolution", *Proc. 31st Annual Conference of the Association for Computational Linguistics* (ACL-93), 62-69, Columbus, Ohio, June 1993

Kehler, Andrew, Stuart Shieber, "Anaphoric dependencies in ellipsis", *Computational Linguistics*, **23**:3, 457-475, June 1997

Kehler, Andrew, "Current theories of centering for pronoun interpretation: a critical evaluation", *Computational Linguistics*, **23**:3, 457-475, June 1997

Kempson, R., "Anaphoric binding, the compositionality requirement and the semantics-pragmatics distinction", C. Jones, P.Sells, eds., *Proc. NELS 14* (North East Linguistic Society), U. Massachusetts, Amherst, GLSA Publications, 1984

Kempson, R., "Definite NPs and context-dependence: a unified theory of anaphora", T. Myers, K. Brown, B. McGonigle, eds., *Reasoning and Discourse Process,* Academic Press, London, 209-239, 1986

Kempson, R., "Anaphora: a unitary account", *Proc. Workshop on Anaphora*, Ofir, Portugal, 1-36, 1990

Kennedy, Christopher, Branimir Boguraev, “Anaphora for everyone: pronominal anaphora resolution without a parser”, *Proc. COLING ‘96: The 16th International Conference on Computational Linguistics*, Copenhagen, Denmark, August 5-9, 1996.

Kennedy, Christopher, Branimir Boguraev, “Anaphora in a wider context: Tracking discourse referents”, W. Wahlster, ed., *12th Eurpean Conference on Artificial Intelligence*, J. Wiley, London, New York, 1996

Kibble, Rodger, “Dynamics of epistemic modality and anaphora”, Bunt, Muskens, Rentier, eds., *Proc. International Workshop on Computational Semantics*, ITK, Tilburg University, The Netherlands, 1994

Kibble, Rodger, “Modal subordination, focus and complement anaphora”, *Proc. Tbilisi Symposium on Language, Logic and Computation*, CSLI, Stanford, Ca.,1995
.
Kibble, Rodger, *Anaphora Resolution in Dynamic Semantics*, Ph.D. Thesis, Centre for Cognitive Science, University of Edinburgh, Scotland, 1997

Kibble, Rodger, “Complement anaphora and dynamic binding”, *Proc. 7th Annual Conference on Semantics and Linguistic Theory* (SALT VII), Stanford University, March 1997

Kibble, Rodger, “Complement anaphora and monotonicity”, *Proc. Formal Grammar Conference*, Aix-en-Provence, France, August 1997

Kleiber, Georges, “Anaphores et pronoms”, *Champs Linguistiques*, Editions Duculot, Louvain-la-Neuve, France, 1994

Krifka, Manfred, “Parametrized sum individuals for plural anaphora”, *Linguistics and Philosophy*, **19**:6, 555-598, December 1996

Kruiff, Geert-Jan M., Ivana Korbayova, “Resolution of direct anaphora in CYAN”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Kuno, S., “Anaphora and discourse principles”, M. Nagao, ed., *Language and Artificial Inelligence*, 87-111, North-Holland, Amsterdam, Holland, 1987

L

Lappin, Shalom, “Pronominal binding and coreference”, *Theoretical Linguistics*, **12**, 241-263, 1985

Lappin, Shalom, Michael McCord, “A syntactic filter on pronominal anaphora in slot grammar”, *Proc. 28th Annual Meeting of the Association for Computational Linguistics (ACL),* 135-142, 1990

Lappin, Shalom, Michael McCord, "Anaphora resolution in slot grammar", *Computational Linguistics,* **16**:4, 197-212, 1990

Lappin, Shalom, Herbert J. Leass, "An algorithm for pronominal anaphora resolution", *Computational Linguistics*, **20**:4, 535-561, 1994

Lasersohn, Peter, "Bare plurals and donkey anaphora", *Natural Language Semantics*, **5**:1, 79-86, 1997

Lasnik, Howard, "Remarks on coreference", *Linguistic Analysis*, **2**:1, 1976

Lasnik, Howard, *Essays on Anaphora, Studies in Natural Language and Linguistic Theory,* Kluwer, Dordrecht, Germany, 1989

Leass, Herbert, Ulrike Schwall, "An anaphora resolution procedure for machine translation", *IWBS Report 172*, IBM Germany Scientific Center, Heidelberg, Germany, 1991

Lebeaux, D., "Anaphoric binding and the definition of PRO", C. Jones, P.Sells, eds., *Proc. North East Linguistic Society (NELS) 14*, University of Massachusetts, Amherst, GLSA Publications, 1984

Lee, H., "Anaphora and the principle of referential autonomy", M. Gonzalez, ed., *Proc. North East Linguistic Society (NELS) 24*, University of Massachusetts, Amherst, GLSA Publications, 1994

Lee, J., K. Stenning, "Anaphora in multimodal discourse", Harry Bunt, Robert-Jan Beun, Tijn Borghuis, eds., *Multimodal Human-Computer Communication*, 250-263, Springer Verlag, Berlin, Germany, 1998

Li, Naicong, David Zubin, "Anaphor resolution in Mandarin", *Proc. Eastern States Conference on Linguistics (ESCOL) '86*, Ohio State University, Columbus, Ohio, CLC Publications, 1987

Lin, Y.Q., R.P. Fawcett, "Anaphoric reference and logical form", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Lockman, A., A.D. Klappholz, "Towards a procedural model of contextual reference resolution", *Discourse Processes*, **3**, 25-71, 1980

Loebner, Sebastian, "Indirect anaphora, types of nouns and cognitve schemata", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Lorenz, Sven, “Presupposition, anaphora and reasoning about change”, *Proc. ECAI*, Vienna, 533-537, 1992

LuperFoy, Susann, “A computational approach to indirect anaphora processing”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Lust, B., T. Clifford, “The 3D study: effects of depth, directionality and distance on children's acquisition of anaphora: an initial report”, J. Pustejovsky, P. Sells, eds., *Proc. North East Linguistic Society (NELS) 12,* Massachusetts Institute of Technology, GLSA Publications, 1982

Lust, B., L. Mangione, “The principle branching direction parameter in first language acquisition of anaphora”, P. Sells & C. Jones, eds., *Proc. North East Linguistic Society (NELS) 13*, Université du Québec à Montréal, GLSA Publications, 1983

M

Maes, Alfons, “Processing abstract anaphora in discourse”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Masuko, M., “Representation and interpretation: a case of anaphora resolution”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Mitkov, Ruslan, H.K. Kim, H.K. Lee, K.S. Choi, “The lexical transfer of pronominal anaphora in Machine Translation: the English-to-Korean case”, *Proc. SEPLN ’94 Conference*, Cordoba, Spain, July 1994

Mitkov, Ruslan, “An integrated model for anaphora resolution”, *Proc. 15th International Conference on Computational Linguistics COLING '94*, Kyoto, Japan, August 5-9, 1994

Mitkov, Ruslan, “A new approach for tracking center”, *Proc. International Conference: New Methods in Language Processing,* UMIST, Manchester, UK, September 13-16, 1994

Mitkov, Ruslan, “Anaphora resolution”, *Natural Language Processing and Machine Translation*, IAI Working Paper, Saarbrücken, 1995 (46 pages)

Mitkov, Ruslan, S.K. Choi, R. Sharp, “Anaphora resolution in Machine Translation”, *Proc. 6th International Conference on Theoretical and Methodological Issues in Machine Translation,* Leuven, Belgium, July 5-7, 1995

Mitkov, Ruslan, Více C.M., "A statistically-based and knowledge-independent approach for anaphora resolution", *Proc. Conference on Formal and Natural Languages,* Tortosa, Spain, September 25-27, 1995

Mitkov, Ruslan, "An uncertainty reasoning approach to anaphora resolution", *Proc. Natural Language Pacific Rim Symposium*, December 4-7, 1995, Seoul, Korea

Mitkov, Ruslan, "Two engines are better than one: generating more power and confidence in the search for the antecedent", *Proc. International Conference on Recent Advances in Natural Language Processing,* September 14-16, 1995, Bulgaria

Mitkov, Ruslan, "Anaphor resolution in Natural Language Processing and Machine Translation", *Proc. International Colloquium on Discourse Anaphora and Anaphora Resolution,* Lancaster, July 17-19, 1996 (keynote speech)

Mitkov Ruslan, "Anaphor resolution: combination of linguistic and statistical approaches", *Proc. International Colloquium on Discourse Anaphora and Anaphora Resolution,* Lancaster, July 17-19, 1996

Mitkov, Ruslan, "Attacking anaphora on all fronts", A. Ramsey, ed., *Artificial Intelligence: Methodology, Systems, Applications*, IOS Press, 1996

Mitkov, Ruslan, "Machine Translation and anaphora", *Machine Translation Review*, No. 4, 1996

Mitkov, Ruslan, "Pronoun resolution: the practical alternative", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Mitkov, Ruslan, "Factors in anaphora resolution: they are not the only things that matter. A case study based on two different approaches", *Proc. ACL '97/EACL '97 (European Chapter of the Association for Computational Linguistics) Workshop on Operational Factors in Practical, Robust Anaphora Resolution*, Madrid, Spain, July 11, 1997

Mitkov, Ruslan, "How far are we from (semi-)automatic annotation of anaphoric links in corpora?", *Proc. ACL '97/EACL '97 Workshop on Operational Factors in Practical, Robust Anaphora Resolution*, Madrid, Spain, July 11, 1997

Mitkov, Ruslan, K.H. Lee, H.G. Kim, K.S. Choi, "English-to-Korean Machine Translation and anaphora resolution", *J. Literary and Linguistics Computing*, **12**:1, 1997

Mitkov, Ruslan, P. Schmidt, "On the complexity of anaphora resolution in Machine Translation", Carlos Martin-Více, ed., *Mathematical linguistics – II*, John Benjamins, 1997

Mitkov, Ruslan, "Anaphora resolution: the state of the art", *COLING-ACL '98 Tutorial Notes*, Montreal, Canada, Aug. 9-14, 1998

Mitkov, Ruslan, “Robust pronoun resolution with limited knowledge”, *Proc. COLING-ACL ’98*, Montreal, Canada, 869-875, Aug. 9-14, 1998

Mitkov, Ruslan, *Evaluating approaches in anaphora resolution*, forthcoming

Mitkov, Ruslan, “Pronoun resolution: the practical alternative”, S. Botley, T. McEnery, eds., *Discourse Anaphora and Anaphor Resolution*, University College London Press, forthcoming

Mitkov, Ruslan, “Two engines are better than one: generating more power and confidence in the search for the antecedent”, Ruslan Mitkov, N. Nicolov, eds., *Recent Advances in Natural Language Processing*, John Benjamins Publishers (forthcoming)

Morgan, G., “Spatial anaphoric mechanisms in British Sign Language”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Moulin, Bernard, “Temporal contexts for discourse representation: an extension of the conceptual graph approach”, *Applied Intelligence*, **7**:3, 227-255, July 1997

Murata, M., *Anaphor Resolution in Japanese Sentences using Surface Expressions and Examples*, Ph.D. Thesis, Kyoto University, Kyoto, Japan, December 1996

Murata, M., M. Nagao, “Indirect reference in Japanese sentences”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Muskens, R., “Anaphora and the logic of change”, J. van Eijck, ed., *Proc. Logics in AI European Workshop JELIA '90*, Amsterdam, The Netherlands, September 1990, *Lecture Notes in Artificial Intelligence* 478, 412-427, Springer Verlag, Berlin, 1991

N

Nakaiawa, Hiroshi, “Discourse comprehension based on activity propagation”, *J. Association of Artificial Intelligence* (Japan), **7**:6, 138-144, Nov. 1992

Nakaiawa, Hiroshi, S. Ikehara, “Zero pronoun resolution in a Japanese-to-English Machine Translation system by using verbal semantic attributes”, *Proc. 3rd Conference on Applied Natural Language Processing* (ANLP'92), Trento, Italy, 1992

Nakaiawa, Hiroshi, A. Yokoo, S. Ikehara, “A system of verbal semantic attributes focused on the syntactic correspondence between Japanese and English”, *Proc. 15th International Conference on Computational Linguistics* (COLING '94), Kyoto, Japan, 1994

Nakaiawa, Hiroshi, S. Shirai, S. Ikehara, T. Kawaoka, "Extrasentential resolution of Japanese zero pronouns using semantic and pragmatic constraint", *Proc. AAAI 1995 Spring Symposium Series: Empirical Methods in Discourse Interpretation and Generation*, 1995

Nakaiawa, Hiroshi, S. Ikehara, "Intrasentential resolution of Japanese zero pronouns in a Machine Translation system using semantic and pragmatic constraints", *Proc. International Conference on Theoretical and Methodological Issues in Machine Translation* (TMI'95), Leuven, Belgium, 1995

Nakaiwa, H., S. Shirai, and S. Ikehara, "Anaphora resolution of Japanese zero pronouns with deictic reference", *16th International Conference on Computational Linguistics* (COLING '96), 812-817, August 1996

Nash-Webber, Bonnie Lynn – see also Webber, Bonnie Lynn

Nash-Webber, Bonnie Lynn, Raymond Reiter, "Anaphora and logical form: on formal meaning representations for natural language", *Proc. 5th IJCAI*, Cambridge, Mass., 1977

Nash-Webber, Bonnie Lynn, "Inference in an approach to discourse anaphora", Brita Warvik, Sanna-Kaisa Tanskanen, Risto Hiltunen, eds., *Proc. North East Linguistic Society (NELS) 8*, University of Massachusetts, Amherst, GLSA Publications, 1978

Nasukawa, T., "Robust method of pronoun resolution using full-text information", *Proc. 15th International Conference on Computational Linguistics* COLING '94, Kyoto, Japan, August 5-9, 1994

Nelken, Rani, Nissim Francez, "Splitting the reference time: temporal anaphora and quantification in DRT", *Proc. 7th Conference of the European Chapter of the Association for Computational Linguistics* (EACL), University College Dublin, Ireland, March 27-31, 1995

Nicol, J., D. Swinney, "The role of structure in coreference assignment during sentence comprehension", *J. Psycholinguistic Research*, 1989

Not, Elena, Zancanaro, M., "Exploiting the discourse structure quantifier for anaphora generation", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

O

Ouhalla, Jamal, "Remarks on the binding properties of Wh-pronouns", *Linguistic Inquiry*, **27**:4, 676-707, Fall 1996

P

Paraboni, Ivandré, Vera Lúcia Strube de Lima, "Possessive pronominal anaphor resolution in Portuguese written texts", *Proc. COLING-ACL '98*, Montreal, Canada, 1010-1014, Aug. 9-14, 1998

Partee, Barbara H., "Nominal and temporal anaphora", *Linguistics and Philosophy*, **7**:3, 243-286, 1984

Passonneau, Rebecca J., *Protocol for Coding Discourse Referential Noun Phrases and Their Antecedents*, Columbia University TR, New York, 1994

Passonneau, Rebecca J.,"Using centering to relax gricean informational constraints on discourse anaphoric noun phrases", *Language and Speech*, **32**:2,3, 229-264, 1996

Paterson, K., A. Sanford, L. Moxey, "Pronominal reference to a quantified noun phrase", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Paterson, K., Edden, R., "Anaphoric reference and scope ambiguity", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Percus, O., "Anaphora and autophagy in Persian and English", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Pirkola, Ari, Kalervo Järvelin, "The effect of anaphor and ellipsis resolution on proximity searching in a text database", *Information Processing & Management*, **32**:2, 199-216, Tiivistelmä, Finland, 1996

Poesio, Massimo, R. Stevenson, J. Hitzeman, "Global focus and pronominalization", D. Jurafsky, T. Regier, eds., *First Conference on Computational Psycholinguistics*, Berkeley, Ca., 1997

Poesio, Massimo, Renata Vieira, and Simone Teufel, "Resolving bridging descriptions in unrestricted text", *Proc. ACL-97 Workshop on Operational Factors in Practical, Robust, Anaphora Resolution For Unrestricted Texts*, Madrid, Spain, 1-6, July 7-11, 1997

Poesio, Massimo, "Cross-speaker anaphora and dialogue acts", H. Rieser, W. Heydrich, eds., *Proc. ESSLLI workshop on Mutual Knowledge, Common Ground and Public Information,* Saarbruecken, Germany, 1998

Pollard, C., I.A. Sag, "Anaphors in English and the scope of binding theory", *Linguistic Inquiry*, **23**:2, 261-303, 1992

Popescu-Belis, Andrei, Isabelle Robba, Gérard Sabah, "Reference resolution beyond coreference: a conceptual frame and its application", *Proc. COLING-ACL '98*, Montreal, Canada, 1046-1052, Aug. 9-14, 1998

Preuß, Susanne, Birte Schmitz, Christa Hauenschild, "Anaphora resolution based on semantic and conceptual knowledge", Susanne Preuß, Birte Schmitz, eds., *Workshop on Text Representation and Domain Modelling - Ideas from Linguistics and AI,* KIT-Report 97, 1-13, Institute for Software and Theoretical CS, Technical University of Berlin, 1992

Preuß, Susanne, Birte Schmitz, Christa Hauenschild, C. Umbach, "Anaphora resolution in machine translation", W. Ramm, ed., *Studies in Machine Translation and Natural Language Processing, Volume 6: Text and Content in Machine Translation: Aspects of Discourse Representation and Discourse Processing,* Office for Official Publications of the European Community, Luxembourg, 1994

Pu, Ming-Ming, "Cognitive constraints, discourse structure and anaphora", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

R

Rasekh, Abbass E., "Discourse anaphora and ambiguity: On the function of context in the retrieving of ambiguous entities", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Rayson, P., A. Wilson, "Anaphora in Market Research Interview Transcipts", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series,**8**, Lancaster University, UK, July 1996

Ravnholt, Ole, "Grammatical cues and 'referential distance' in retrieval of antecedents in discourse", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Ravnholt, Ole, "The role of lexical specifications in the retrieval of direct and indirect antecedents", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Reinhart, Tanya, *The syntactic domain of anaphora*, Ph.D Thesis, Department of Foreign Literature and Linguistics, MIT, Cambridge, MA, 1976

Reinhart, Tanya, *Anaphora and Semantic Interpretation*, Croom Helm, London, also U. Chicago Press, Chicago, 1983

Reinhart, Tanya, “Co-reference and bound anaphora: a restatement of the anaphora question”, *Linguistics and Philosophy*, **6**, 47-88. 1983

Rich, Elaine, Susann LuperFoy, “An architecture for anaphora resolution”, *Proc. Associaton for Computational Linguistics (ACL) Second Conference on Applied Natural Language Processing,* Austin, Texas, 18-24, February 9-12, 1988

Rico Pérez, C., “Estudio de la incidencia de differentes fuentes de la información en el establecimiento de relaciones anafóricas”, *Bulletín de la Sociedad Española para el Procesamiento del Lenguaje Natural,* No. 14, March, 1994

Rico Pérez , C., “Resolución de la anáfora discursiva mediante una estrategia de inspiración vectoral”, *Proc. SEPLN '94 Conference*, Cordoba, Spain, July 20-22,1994

Roberts, C., “Modal subordination and pronominal anaphora in discourse”, *Linguistics and Philosophy*, **12**, 683-721, 1989

Rocha, Marco, *Multimedia Interfaces and Anaphora Resolution*, Isle of Thorns Annual Workshop, University of Sussex at Brighton, England, 1995

Rocha, Marco A.E., “Anaphoric noun phrases of low semantic content and their discourse-constructed antecedents”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Rocha, Marco A.E., “A corpus-based study of anaphora in English and Portuguese”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Rolbert, M., *Résolution de formes pronominales dans l'interface d'interogation d'une base de données*, Ph.D. Thesis, Faculté des sciences de Luminy, France, 1989

S

Saeboe, Kjell Johan, “Anaphoric presuppositions and zero anaphora”, *Linguistics and Philosophy,* **19**:2, 187-209, April 1996

Safir, Ken, “Semantic atoms of anaphora”, *Natural Language and Linguistic Theory,* **14**:3, 545-589, August 1996

Saggion, H., Ar. Carvalho, “Anaphora resolution in a machine translation system”, *Proceedings of the International Conference: Machine Translation, 10 Years On,* Cranfield, UK, November 12-14, 1994

Saggion, H., A. Carvalho, "Definite anaphora in Portuguese abstracts", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Sakakibara, Sonoko, "Pragmatics or syntax? The case of Japanese reflexive pronouns", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Sandum, Gabriel, "On the theory of anaphora: dynamic predicate logic vs. game-theoretical semantics", *Linguistics and Philosophy*, **20**:2, 147-174, April 1997

Sanford, Tony, Simon Garrod, "What, when and how: Questions of immediacy in anaphoric reference and resolution", *Language and Cognitive Processes*, **4**, 235-262, 1989

Sanford, Tony, "Anaphora", M.E. Eysenck, ed., *Dictionary of Cognitive Psychology,* Basil Blackwell, Oxford, 1990

Sanford, Tony, F. Lockhart, "Description types and method of conjoining as factors influencing plural anaphora: A continuation study of focus", *Journal of Semantics*, **7**, 365-378, 1990

Schiehlen, Michael, "Disambiguation of Underspecified Discourse Representation Structures under Anaphoric Constraints", *Proc. Second International Workshop on Computational Semantics*, Tilburg, The Netherlands, January, 1997

Schmitz, Birte, Susanne Preuß, Christa Hauenschild, "Textrepräsentation und Hintergrundwissen für die Anaphernresolution im Maschinellen Übersetzungssystem KIT-FAST", M. Kohrt, Ch. Küper (eds.), *Probleme der Übersetzungswissenschaft*, Working Papers in Linguistics, Department for Linguistics, Technical University of Berlin 1991, 39-81, and in KIT-Report 93, Institute for Software and Theoretical CS, Technical University of Berlin, Germany, 1992

Schuster, E., "Anaphoric reference to events and actions: A representation and its advantages", *Proc. 12th COLING Conference*, 602-607, Budapest, Hungary, 1988

Schwarz, Monika, "Indirect anaphora in text: linguistic and cognitive constraints", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Sells, P., "Coreference and bound anaphora: a restatement of facts", S. Berman, J.W. Choe, J. McDonough, eds., *Proc. North East Linguistic Society (NELS) 16*, McGill University, GLSA Publications, 1986

Seuren, P.A.M., "Anaphora resolution", T. Myers, K. Brown, B. McGonigle, eds., *Reasoning and Discourse Process*, Academic Press, London, 187-207, 1986

Shokouhi, H., "Anaphoric relations in conversation", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Sidner, Candace L., "Towards a computational theory of definite anaphora comprehension in English discourse", *M.I.T. Artificial Intelligence Laboratory Technical Report AI-TR-537*, 1979

Sidner, Candace L., "Disambiguation references and interpreting sentence purpose in discourse", Winston, Brown, eds., *Artificial Intelligence: an M.I.T. Perspective*, **1**, M.I.T. Press, Cambridge, MA, 231-252, 1979

Sidner, Candace L., "Focusing for interpretation of pronouns", *American J. Computational Linguistics,* **7**, 217-231, 1981

Sidner, Candace L., "Focusing in the comprehension of definite anaphora", Michael Brady, Robert Berwick, eds., *Computational Models of Discourse*, 267-330, MIT Press, Cambridge, 1983

Sidner, Candace L., "Focusing in the comprehension of definite anaphora", Barbara Grosz, K. Sparck-Jones, Bonnie Webber, eds., *Readings in Natural Language Processing*, 363-394, Morgan Kaufmann, Los Altos, Ca., 1986 1986

Smyth, R., C. Chambers, "Parallelism effects on pronoun resolution in discourse", *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC '96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Srivinas, B., B. Baldwin, "Exploiting super tag representation for fast coreference resolution", *Proc. International Conference on Natural Language Processing and Industrial Applications,* Moncton, Canada, June 4-6, 1996

Stabler, E., *The Logical Approach to Syntax*, MIT Press, Cambridge, 1992

Stalnaker, Robert, "On the representation of context", *J. Logic, Language and Information*, **7**:1, 3-19, January 1998

Stevenson, Rosemary, R.A. Crawley, G. Wilson, D. Kleinman, "Thematic roles and pronoun comprehension", *Proc. 12th Annual Conference of the Cognitive Science Society*, 534-541, 1990

Stevenson, Rosemary, A.W.R. Nelson, K. Stenning, "Strategies in pronoun comprehension", *Proc. 15th Annual Conference of the Cognitive Science Society*, 976-981, 1993

Stevenson, Rosemary, A. Urbanowicz, “Structural focusing, thematic role focusing and the comprehension of pronouns”, *Proc. 17th Annual Conference of the Cognitive Science Society,* 328-332, Pittsburgh, PA, 1995

Stevenson, Rosemary, A. Urbanowicz, “The effects of sentence subject, initial mention and pragmatic plausibility on the accessibility of a pronoun's antecedents”, Human Communication Research Center, Universities of Edinburgh and Glasgow, Scotland, Research Paper HCRC/RC-58, 1995

Strand, Kjetil, “An application of inside-out functional uncertainty to Norwegian anaphors”, Kimmo Koskenniemi, ed., *Proc. 10th Nordic Conference on Computational Linguistics*, Department of General Linguistics, University of Helsinki, Helsinki, Finland, May 30-31, 1995

Strube, Michael, Udo Hahn, “ParseTalk about sentence- and text-level anaphora”, *Proc. 7th Conference of the European Chapter of the Association for Computational Linguistics* (EACL), University College Dublin, Ireland, March 27-31, 1995

Stuckardt, Roland, “An interdependency-sensitive approach to anaphor resolution”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, 400-413, Lancaster University, July 1996

Stuckardt, Roland, “Anaphor resolution and the scope of syntactic constraints”, *Proc. 16th International Conference on Computational Linguistics* (COLING), **2**, 932-937, Copenhagen, August 1996

Stuckardt, Roland, *Algorithmische Interpretation anaphorischer Ausdruecke*, Arbeitspapiere der GMD, No. 1045, GMD-Forschungszentrum Informationstechnik GmbH, D-53754 Sankt Augustin, January 1997

Stuckardt, Roland, “Resolving Anaphoric References on Deficient Syntactic Descriptions”, *Proc. ACL '97 /EACL '97 Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, 30-37, Madrid, Spain, July 11, 1997

Stys, M., S. Zemke, “Incorporating discourse aspects in English - Polish MT: towards robust implementation”, *Proc. International Conference on Recent Advances in Natural Language Processing,* Tzigov Chark, Bulgaria, September 14-16, 1995

T

Tanaka, I.., “Using annotated corpora to investigate pronoun resolution”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Tannenhaus, M. K., G. N. Carlson, “Processing deep and surface anaphors”, S. Berman, J-W. Choe, J. McDonough, eds., *Proc. North East Linguistic Society (NELS) 15,* Brown University, GLSA Publications, 1985

ter Meulen, A., “Aspectual adverbs and their anaphoric presuppositions”, L. Gabriele, D. Hardison, & R. Westmoreland , eds., *Proc. FLSM VI: Papers from the Sixth Annual Meeting of the Formal Linguistics Society of Mid-America*, **2**, Indiana University, Bloomington, IN, IULC Publications, May 1995

Tin, E., V. Akman, “Anaphora in Turkish: A preliminary implementation”, *6th International Conference on Turkish Linguistics*, Eskisehir, Turkey, 1992

Tin, E., V. Akman, “Situated processing of pronominal anaphora”, H. Trost, ed., *Proc. KONVENS ‘94 Conference on Natural Language Research*, 369-378, Vienna, September 28-30, 1994

Tsurusaki, T., “Bach-Peters paradox and two modes of pronominal anaphora”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Turan, U.D., “Zero object arguments and referentiality in Turkish”, *7th International Conference on Turkish Linguistics,* Institute of Oriental Studies, Turkology, Johannes Gutenberg Univ., Mainz, Germany, 1994

Tutin, A., E. Viegas, “The resolution and generation of anaphoric definite expressions”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Tyler, Lorrain K., William Marslen-Wilson, “Resolution of discourse anaphora”, *Text*, **2**, 1982

U

Uehara, S., “Anaphoric pronouns in English and their counterparts in Japanese”, *Proc. International Colloquium on Discourse Anaphora and Anaphora Resolution*, Lancaster, UK, July 17-19, 1996

V

Verkuyl, H. J., C. F. M. Vermeulen, “Shifting perspectives in discourse”, *Linguistics and Philosophy,* **19**:5, 503-526, October 1996

Verschuur, Leon, “Semantic agreement of anaphors and extended type theory”, *Proc. ESCOL ’93* (Eastern States Conference on Linguistics), CLC Publications, Cornell University, Ithaca, NY, 1994

Von Heusinger, Klaus, "Anaphorisches varieté", *Tagung Sinn und Bedeutung*, U. Stuttgart, Tübingen, Germany, Dec. 19-21, 1996

W

Wada, H., "Discourse processing in MT: problems in pronominal translation", *Proc. 13th International Conference on Computational Linguistics* (COLING '90), Helsinki, Finland, 1990

Walker, Marilyn A., "Evaluating discourse processing algorithms", *Proc. 27th Annual Meeting of the Association for Computational Linguistics*, 251-261, 1989

Walker, Marilyn A., Masayo Iida, Sharon Cote, "Centering in Japanese discourse", *Proc. 13th International Conference on Computational Linguistics* (COLING '90), 1-8, Helsinki, Finland, 1990

Walker, Marilyn A., Masayo Iida, Sharon Cote, "Japanese discourse and the process of centering", *Computational Linguistics*, **20**:2, 1994

Walker, Marilyn A., "Centering, anaphora resolution, and discourse structure", Marilyn A. Walker, Ellen F. Prince, A. Joshi, eds., *Centering in Discourse,* Oxford University Press, 1997

Walker, Marilyn A., Ellen F. Prince, A. Joshi, eds., *Centering in Discourse,* Oxford University Press, 1997

Webber, Bonnie Lynn – see also Nash-Webber, Bonnie L.

Webber, Bonnie Lynn, Raymond Reiter, "Anaphora and logical form: on formal meaning representations for natural language", *Proc. IJCAI '77*, 121-131, 1977

Webber, Bonnie Lynn, *A Formal Approach to Discourse Anaphora*, Ph.D. Thesis, Harvard University, 1978

Webber, Bonnie Lynn, *A Formal Approach to Discourse Anaphora*, Garland Publishing, London, 1979

Webber, Bonnie Lynn, "Discourse deixis: Reference to discourse segment", *Proc. 26th Annual Meeting of the Association for Computational Linguistics,* 113-122, June 1988

Webber, Bonnie Lynn, "Tense as discourse anaphor", *Computational Linguistics*, **14**:2, 61-73, 1988

Webber, Bonnie Lynn, "Structure and ostension in the interpretation of discourse deixis", *Language and Cognitive Processes*, **6**:2, 107-135, 1991

Wilks, Y.A., “Preference semantics”, E. Keenan, ed., *The Formal Semantics of Natural Language*, Cambridge University Press, 1975

Williams, Edwin, “Blocking and anaphora”, *Linguistic Inquiry*, **28**:4, 577-628, Fall 1997

Williams, S., “Anaphoric reference resolution in a telephone-based spoken language system for accessing email”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Williams, S., K. Preston, M. Harvey, “Rule-based reference resolution resolution for unrestricted text using part-of-speech tagging and anaphor/antecedent noun-phrase parsing”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Winograd, T., *Understanding Natural Language*, Academic Press, New York, 1972

Y

Yabushita, Katsuhiko, “'Cart-Sentence' anaphora: Is Kamp/Heim's treatment of pronouns really a unified one?”, *Proc. ESCOL ’88* (Eastern States Conference on Linguistics), Ohio State University, Columbus, Ohio, CLC Publications, 1989

Yamada,I., T. Yamamura, Y. Sagawa, N. Ohnishi, N. Sugie, “On analysis of the characteristic of unknown words and examination of the available method for inferring meanings”, *Tr. IPSJ*, **35**:12, 2725-2733, December 1994 (in Japanese)

Yamamura, T., N. Ohnishi, N. Sugie, “A system resolving anaphora reference problems in Japanese”, *Tr. IEICE*, **J73-D-II**:6, 887-896, June 1990 (in Japanese)

Yamamura, T., N. Ohnishi, N. Sugie, “A classification scheme of anaphora in Japanese demonstrative pronouns”, *Tr. IEICE*, **J75-D-II**:2, 371-378, Feb. 1992

Yamamura, T., N. Ohnishi, N. Sugie, “A method of anaphora resolution based on concept of observation”, *Tr. IEICE*, **J77-D-II**:1, 162-169, January 1994 (in Japanese)

Yamamura, T., N. Ohnishi, N. Sugie, “A method for anaphora resolution based on the concept of observation”, *Systems and Computers in Japan*, **26**:1, 40-49, Scripta Technica Inc., 1995

Yang, D.W., “On anaphor movement”, J. Carter, R.M. Dechaine, eds., *Proc. North East Linguistic Society (NELS) 19,* Cornell University, Ithaca, NY, GLSA Publications, 1989

Yeh, Ching-Long, *Generation of Anaphors in Chinese*, Ph.D. Thesis, Edinburgh University, Edinburgh, Scotland, 1995

Yeh, Ching-Long, Chris Mellish, “An empirical study on the generation of anaphora in Chinese”, *Computational Linguistics*, **23**:1, 169-190, March 1997

Z

Ziv, Yael, “Inferred antecedents and epithets: clues in anaphor resolution”, *Approaches to Discourse Anaphora: Proc. Discourse Anaphora and Anaphor Resolution Colloquium* (DAARC ‘96), UCREL Technical Papers Series, **8**, Lancaster University, UK, July 1996

Zuri, L., K. F. McCoy, “Focusing and pronoun resolution in particular kinds of complex sentences”, *Workshop on Centering Theory in Naturally Occurring Discourses,* University of Pennsylvania, May 1993