ANAPHORA RESOLUTION: THE STATE OF THE ART

Ruslan Mitkov

School of Languages and European Studies
University of Wolverhampton
Stafford Street
Wolverhampton WV1 1SB
United Kingdom
Email R.Mitkov@wlv.ac.uk
http://www.wlv.ac.uk/~le1825/

1. Introduction

Anaphora resolution is a complicated problem in Natural Language Processing and has attracted the attention of many researchers. The approaches developed - traditional (from purely syntactic ones to highly semantic and pragmatic ones), alternative (statistic, uncertainty-reasoning etc.) or knowledge-poor, offer only approximate solutions.

The paper is an introduction to anaphora resolution offering a brief survey of the major works in the field.

1.1 Basic notions and terminology

The etymology of the term "anaphora" goes back to Ancient Greek with “anaphora” (αναφορα) being a compound word consisting of the separate words ανα – *back, upstream, back in an upward direction* and φορα - *the act of carrying* and denoted *the act of carrying back upstream*. For Computational Linguists embarking upon research in the field of anaphor resolution, I strongly recommend as a primer Graham Hirst's book "Anaphora in natural language understanding" (Hirst 1981) which may seem a bit dated in that it does not include developments in the 80's and the 90's, but which provides an excellent survey of the theoretical work on anaphora and of the early computational approaches and is still very useful reading.

Various definitions of anaphora have been put forward, but I am tempted to paraphrase the classical definition given by Halliday and Hasan (Halliday & Hasan 1976) which is based on the notion of cohesion: anaphora is cohesion (presupposition) which points back to some previous[1] item.

[1]We shall not discuss cataphora which is the case when the "anaphor" precedes the antecedent (e.g. "Because she was going to the post office, Julie was asked to post a small parcel")

The "pointing back" (reference) is called an *anaphor* and the entity to which it refers is its *antecedent*. The process of determining the antecedent of an anaphor is called *anaphora resolution*. Usually, both the antecedent and the anaphor are used as referring expressions and having the same referent[2] in the real world, they are termed *coreferential*.

Example (Huddleston 1984):

The Empress hasn't arrived yet but she should be here any minute

In this example, the pronoun "she" is the anaphor (for classification of anaphors, see below) and "the Empress" is the antecedent. Please note that the antecedent is not the noun "Empress" but the noun phrase "the Empress".

There may be cases when the anaphor and more than one of the preceding (or following) entities (usually noun phrases) have the same referent and are therefore pairwise coreferential, thus forming a coreferential chain. In such a case, we regard each of the preceding entities which are coreferential with the anaphor(s) as a legitimate antecedent. Therefore, in such cases the task of anaphora resolution is considered successful, if any of the preceding entities in the coreferential chain is identified as an antecedent. Our paper will discuss the task of anaphora resolution only and not coreference resolution (except for briefly mentioning it in section 4.2). For more on coreference resolution, I suggest the reader consult the MUC (Message Understating Conference) Proceedings in which coreference resolution is extensively covered.

1.2 Types of anaphora

There are various types of anaphora (Hirst 1981), but we will shall briefly outline those that are thought to be the three most widespread types in the Computational Linguistics literature.

- Pronominal anaphora

The most widespread type of anaphora is the pronominal anaphora which is realised by anaphoric pronouns.

Example:

Computational Linguists from many different countries attended the tutorial.
They took extensive notes.

It should be pointed out that not all pronouns in English are anaphoric. For instance, "it" can often be non-anaphoric such as in the case of the previous sentence. Other examples of non-anaphoric “it” include expressions such as "It is important", "It is necessary", "It has to be taken into account". A non-anaphoric "it" is termed pleonastic (Lappin & Leass 1994).

---

[2]The relation between the anaphor and the antecedent is not to be confused with that between the anaphor and its referent; in the example below the referent is "the Empress" as a person in the real word whereas the antecedent is "the Empress" as a linguistic form

• Definite noun phrase anaphora

Typical cases of definite noun phrase anaphora is when the antecedent is referred by a definite noun phrase representing either same concept (repetition) or semantically close concepts (e.g. synonyms, superordinates).

Example:

> Computational Linguists from many different countries attended the tutorial. The participants found it hard to cope with the speed of the presentation.

• One-anaphora

One-anaphora is the case when the anaphoric expression is realised by a "one" noun phrase. Example:

> If you cannot attend a tutorial in the morning, you can go for an afternoon one.

Finally, we distinguish *intrasentential* anaphors (referring to an antecedent which is in the same sentence as the anaphor) and *intersentential* anaphors (referring to an antecedent which is in a different sentence from that of the anaphor).

1.3 The process of anaphora resolution

Most of the anaphora resolution systems deal with resolution of anaphors which have noun phrases as their antecedents because identifying anaphors which have verb phrases, clauses, sentences or even paragraphs/discourse segments as antecedents, is a more complicated task. Typically, all noun phrases (NPs) preceding an anaphor are initially regarded as potential candidates for antecedents. Usually, a search scope has to be identified: most approaches look for NPs in the current and preceding sentence. However, an "ideal" anaphora resolution system should extend its scope of search: antecedents which are 17 sentences away from the anaphor have already been reported (Mitkov 1995a)!

Assuming that the scope of search for a specific approach has already been specified, the NPs preceding the anaphor within that scope are identified as candidates for antecedents and a number of anaphora resolution factors are employed to track down the correct antecedent.

Approaches to anaphora resolution usually rely on a set of "anaphora resolution factors". Factors used frequently in the resolution process include gender and number agreement, c-command constraints, semantic consistency, syntactic parallelism, semantic parallelism, salience, proximity etc. These factors can be "eliminating" i.e. discounting certain noun phrases from the set of possible candidates (such as gender and number constraints[3], c-command constraints, semantic consistency) or "preferential", giving more preference to certain candidates and less to others (such as parallelism, salience). Computational linguistics literature uses diverse terminology for these - for example E. Rich and S. LuperFoy (Rich & LuperFoy 1988) refer to the "eliminating"

---

[3] In English antecedents usually agree in gender and number with the anaphors, but this is not always the case in other languages

factors as "constraints", and to the preferential ones as "proposers", whereas Carbonell and Brown (Carbonell & Brown 1988) use the terms "constraints" and "preferences". Other authors argue that all factors should be regarded as preferential, giving higher preference to more restrictive factors and lower - to less "absolute" ones, calling them simply "factors" (Preuß et al. 1994), "attributes" (Rico Pérez 1994), "symptoms" (Mitkov 1995b) or "indicators" (Mitkov 1996a, 1998b).

The division of factors into constraints and preferences has led to distinguishing between constraint-based and preferences-based architectures in anaphora resolution (Mitkov 1997b).

1.3.1 *Constraints*

Several constraints will be outlined and illustrated by examples. Coreferential items are given the same index.

- Gender and number agreement

This constraint requires that anaphors and their antecedents must[4] agree in number and gender.

Example:

Jane$_i$ told Philip$_k$ and his friends$_m$ that she$_i$ was in love.

- Syntactic binding theories' constraints

Results in Government and Binding Theory (GB)[5] and Lexical Functional Grammar have provided useful constraints on the anaphors and their antecedents which have been successfully used in anaphor resolution. For instance, various GB c-command restrictions have been formulated in (Ingria & Stallard 1989) for eliminating unacceptable candidates when searching for the antecedent:

(a) A non-pronominal NP cannot overlap in reference with any NP that c-commands it.

He$_i$ told them about John$_j$.

(b) The antecedent of a bound anaphor must c-command it.

John$_i$ likes pictures of himself$_i$.

(c) A personal pronoun cannot overlap in reference with an NP that c-commands it.

John$_i$ told Bill$_j$ about him$_k$.

[4] There are certain collective nouns in English which do not agree in number with their antecedents (e.g. "government", "team", "parliament" etc. can be referred to by "they"; equally some plural nouns such as "data" can be referred to by "it") and should be exempted from the agreement test

[5] For more on c-commands see L. Haegeman's book (Haegeman, 1994)

- Semantic consistency

This constraint stipulates that if satisfied by the anaphor, semantic consistency constraints must be satisfied also by its antecedent.

> Vincent removed the diskette from the computer$_i$ and then disconnected it$_i$.
> Vincent removed the diskette$_i$ from the computer and then copied it$_i$.

1.3.2 *Preferences*

Preferences, as opposed to constraints, are not obligatory conditions and therefore do not always hold. We shall illustrate three preferences: syntactic parallelism, semantic parallelism and center of attention.

- Syntactic parallelism

Syntactic parallelism could be quite helpful when other constraints or preferences are not in a position to propose an unambiguous antecedent. This preference is given to NPs with the same syntactic function as the anaphor.

> The programmer$_i$ successfully combined Prolog$_j$ with C, but he$_i$ had combined it$_j$ with Pascal last time.
> The programmer$_i$ successfully combined Prolog with C$_j$, but he$_i$ had combined Pascal with it$_j$ last time.

- Semantic parallelism

This is a useful (and stronger than syntactic parallelism) preference but only systems which can automatically identify semantic roles, can employ it. It says that NPs which have the same semantic role as the anaphor, are favoured.

> Vincent gave the diskette to Sody$_i$. Kim also gave him$_i$ a letter.
> Vincent$_i$ gave the diskette to Sody. He$_i$ also gave Kim a letter.

- Centering

Although the syntactic and semantic criteria for the selection of an antecedent are very strong, they are not always sufficient to distinguish between a set of possible candidates. Moreover, they serve more as filters to eliminate unsuitable candidates than as proposers of the most likely candidate. In the case of antecedent ambiguity, it is the most salient element among the candidates for antecedent which is usually the front-runner. This most salient element is referred to in computational linguistics as focus (e.g. (Sidner 1979) or center[6] e.g. (Grosz et al. 83) though the terminology can be much more diverse (Hirst 1981; Mitkov 1995a).

[6] Center and focus are close, but not identical concepts. We refer the reader to (Grosz et al. 1995)

For instance, neither machines, nor humans, would be able to resolve the anaphoric pronoun "it" in the sentence

> Jenny put the cup on the plate and broke it.

However, if this sentence is part of a discourse segment[7] which makes it possible to determine the most salient element, the situation is different:

> Jenny went window shopping yesterday and spotted a nice cup. She wanted to buy it, but she had no money with her. Nevertheless, she knew she would be shopping the following day, so she would be able to buy the cup then. The following day, she went to the shop and bought the coveted cup. However, once back home and in her kitchen, she put the cup on a plate and broke it...

In this discourse segment, "the cup" is the most salient entity and is the center of attention throughout the discourse segment.

It is now clear that very often when two or more candidates "compete" for the antecedent, the task of resolving the anaphor is shifted to the task of tracking down the center/focus of the sentence (clause). Various methods have already been proposed to center/focus tracking (e.g. Brennan et al. 1987; Dahl & Ball 1990; Mitkov 1994b; Sidner 1986; Stys & Zemke 1995; Walker et al. 92).

However useful the term center (or focus) can be for anaphora resolution, we should point out that it has suffered from two inconveniences: its intuitive nature and the use of different terms to describe concepts which either seem to be very close to "center" or even could be considered practically identical (e.g. focus, topic, theme - for further details please see (Hirst 1981) and (Mitkov 1995a).

1.3.3 *Computational strategies*

While a number of approaches use a similar set of factors, the "computational strategies" for the application of these factors may differ (here, the term "computational strategy" refers to the way antecedents are computed, tracked down, i.e. the algorithm, formula for assigning antecedents rather than computational issues related to programming languages, complexity etc.). Some approaches incorporate a traditional model which discounts unlikely candidates until a minimal set of plausible candidates is obtained (and then make use of center or focus, for instance), whereas others compute the most likely candidate on the basis of statistical or AI techniques/models. This observation led us to categorise the approaches to anaphora resolution as "traditional" or "alternative" (see section 3).

2. Early work on anaphora resolution

In this survey I shall not present the early works on anaphor resolution comprehensively since they are very well covered in (Hirst 1981): I shall mention only Bobrow's program STUDENT, Winograd's SHRDLU, Wilk's preference semantics approach and Hobb’s naïve algorithm.

---

[7] For definition of discourse segment see (Allen, 1995), Chapter 14

2.1 STUDENT (Bobrow 1964)

One of the earliest attempts to resolve anaphors by a computer program is reported in STUDENT (Bobrow 1964), a high-school algebra problem answering system. It has a few limited heuristics for resolving anaphors and more particularly anaphora paraphrases and incomplete repetitions. For example, the system can successfully track down the antecedent in the following example using pattern matching:

> The number of soldiers the Russians have is half the number of guns they have. The number of guns is 7000. What is the number of soldiers they have?

However, these simple heuristics are not as intelligent as they might seem since the sentence is not even parsed in any real sense. The following references to 'sailors' for example, could not be matched up:

> The number of soldiers the Russians have is twice the number of sailors they have. The number of soldiers is 7000. How many sailors do the Russians have?

2.2 SHRDLU (Winograd 1972)

Winograd (Winograd 1972) was the first to develop procedures for pronoun resolution in his SHRDLU system. He checks previous noun phrases for possible antecedents and does not consider the first likely candidate but examines all the possibilities in the preceding text and rates plausibility on the basis of syntactic position (subject is favoured over object and both are favoured over the object of a preposition). In addition, "focus" elements are favoured, the focus being determined from the answers to wh-questions and from indefinite noun phrases in yes-no questions.

2.3 Preference semantics (Wilks 1973)

Wilks (Wilks 1973; 1975) describes an English to French translation system which uses 4 levels of pronominal anaphor resolution depending on the type of anaphor and the mechanism needed to resolve it. The lowest level, referred to as type "A" anaphora, uses only knowledge of individual lexeme meanings successfully solving cases such as:

> Give the bananas to the monkeys although they are not ripe, because they are very hungry.

Each "they" is interpreted correctly using the knowledge that the monkeys, being animate, are likely to be hungry and bananas, being fruit, are likely to be ripe.

If a word meaning fails to find a unique antecedent for the pronoun, inference methods for type "B" anaphors - those that need analytic inference - or type "C" anaphors - those that require inference using real-world knowledge beyond the simple word meanings - are brought in. If the anaphor is still unresolved after all this, "focus of attention" rules attempt to find the topic of the sentence to use as the antecedent.

2.4 Taking syntax on board: Hobbs' algorithm

Syntax methods are used usually as filters to discount unacceptable candidates rather than to propose antecedents. Syntax-oriented methods alone are not enough, though they play an important role in anaphor resolution.

One of the first (and best) methods relying mostly on syntactical knowledge is Hobbs' algorithm (Hobbs 1976; 1978). His algorithm is based on various syntactic constraints on pronominalisation which are used to search the tree. The search is done in an optimal order in such a way that the NP upon which it terminates is regarded as the probable antecedent of the pronoun at which the algorithm starts.

Hobbs tests his algorithm on texts from an archaeology book, an Arthur Hailey novel and a copy of Newsweek. He reports (Hobbs 1976) that the algorithm worked 88 per cent of the time, and 92 percent when augmented with simple selectional constraints.

Although Hobbs' algorithm has been regarded as inadequate (Hirst 1981), it was quite a success for its time. Recent comparison with more sophisticated methods (Baldwin 1997; Walker 1989) shows it could be attractive even from the point of view of today's results. Moreover, as Hobbs suggests, his algorithm may still be useful, since it is computationally cheap compared to any semantic method for pronoun resolution.

3. Recent developments in anaphora resolution

After considerable and successful initial research in anaphor resolution (Hobbs 1976; Sidner 1979; Webber 1979, Wilks 1975) and after years of relative silence in the early eighties, the late 80’s and early 90’s saw a revival of interest with various projects reported. We will summarise the most well-known works by grouping them into "traditional" and "alternative" approaches. As I have already explained (subsection 1.3.3), the basis for this classification is the computational strategy used. We term "traditional" those classic approaches that integrate knowledge sources/factors that discount unlikely candidates until a minimal set of plausible candidates is obtained (then make use of center or focus, or other preference). We regard as "alternative" approaches that compute the most likely candidate on the basis of statistical or AI techniques/models.

3.1 Traditional approaches

3.1.1 D. Carter's shallow processing approach

Carter reports in his PhD thesis (Carter 1986) and later in his book (Carter 1987) on a "shallow processing" approach which exploits knowledge of syntax, semantics and local focusing as heavily as possible without relying on the presence of a large amount of world or domain knowledge which, as he comments, is notoriously hard to process accurately.

His ideas are implemented in a program which resolves anaphoric and other linguistic ambiguities in simple English stories and generates sentence-by-sentence paraphrases that show which interpretations have been selected. The program combines and develops several existing theories, most notably Sidner's (Sidner 1979) theory of local focusing and Wilks' (Wilks 1975) "preference semantics" theory of semantics as well as common sense inference.

Carter modifies Sidner's framework and extends it to allow focus-based processing to interact more flexibly with processing being based on other types of knowledge. Wilks' treatment of common sense inference is extended to incorporate a wider range of types of inference without jeopardising its uniformity and simplicity. Wilks' primitive-based formalism for word sense meanings was further developed by Carter.

3.1.2 E. Rich and S. LuperFoy's distributed architecture

E. Rich and S. LuperFoy (Rich & LuperFoy 1988) describe a pronominal anaphor resolution system consisting of a loosely coupled set of modules ("distributed architecture") which handle recency, number agreement, gender agreement, animacy, disjoint reference, semantic consistency, global focus, cataphora and logical accessibility. The authors maintain that each of these modules (referred to as "constraint source" (CS)) represents an implementation of one of the partial theories of anaphor and each of them imposes a set of constraints on the choice of an antecedent.

The constraint sources themselves are composed of a set of four functions and these component functions are called at different times during the process of anaphor resolution. The four functions are modeller (maintains the CS's local model of the ongoing discourse), constraint "poster" ("posts" constraints that describe interactions among anaphora within a sentence), proposer (takes as input an anaphoric reference and returns as its output a list of candidate antecedents, each with an associated score) and evaluator (takes as input an anaphor and a candidate antecedent for that anaphor).

The final selection of an antecedent from among the set of candidates for antecedents depends on the combined score that is attached to each candidate as a result of the examination of the candidate by the entire set of constraint sources. The scoring procedure used is such that each CS provides both a score and a confidence measure. The score is a number in the range 0 to 1 and the function that combines a set of n (score, confidence) pairs is

$$\text{running score} = \left[\sum_{i=1}^{n} \text{score}(i)\text{confidence}(i)\right] / \sum_{i=1}^{n} \text{confidence}(i)$$

3.1.3 J. Carbonell and R. Brown's multi-strategy approach

J. Carbonell and R. Brown report on a multi-strategy approach to anaphor resolution (Carbonell & Brown 1988) on the assumption that anaphor resolution may be best accomplished through the combination of a set of strategies rather than by a single monolithic method. They concentrate on resolving intersentential anaphora, considering this type of anaphora to be more frequent and more crucial in designing interactive natural language interfaces.

The authors propose a general framework for anaphor resolution based on the integration of multiple knowledge sources: sentential syntax, case-frame semantics, dialogue structure and general world knowledge. Their approach is based on a set of constraints and preferences. The constraints are local anaphor constraints (which correspond to agreement constraints) and precondition/postcondition constraints. The latter have to do with real-world knowledge and pragmatics, e.g. in the example "John gave Tom an apple. He ate the apple", "he" refers to Tom, as John no longer has the apple. The postcondition on "give" is that the actor no longer has the object being given by him.

The preferences are case-role persistence preference, semantic alignment preference, syntactic parallelism preference and syntactic topicalisation preference. Case-role persistence preference gives priority to candidates for antecedents which fill the same case role as the anaphor, whereas semantic alignment preference is illustrated by the examples "Mary drove from the park to the club. Peter went there too" (there=club) and "Mary drove from the park to the club. Peter left there too" (in this example, because of the semantics of the verb "to leave", "there" aligns semantically with "park" and not with "club").

The resolution method works by applying the constraints first to reduce the number of candidate antecedents for the anaphor in question. Then, the preferences are applied to each of the remaining candidates. If more than one preference applies, and each suggests different candidate antecedents for the anaphor in question, all of which have passed the constraint tests, then the anaphor is considered to have a truly ambiguous antecedent. Regarding evaluation, the authors mention that manual simulations of 70 examples of a slightly different method have yielded 49 unique resolutions, 17 conflicting possibilities, and 4 anomalous cases (the majority of the 17 multiple-referent cases have been judged to be ambiguous by human readers).

3.1.4 C. Rico Pérez' scalar product coordinating approach

Celia Rico Pérez (Rico Pérez 1994) proposes a method for "coordinating" morphological, syntactical, semantic and pragmatic information by using the scalar product as a measure of closeness or distance between anaphors and candidates for antecedents. Her approach which regards the anaphors as well as the discourse units as vectors, can be summarised as follows:

1. A set of anaphoric attributes for both the anaphoric expressions and the discourse entities is defined. The set of anaphoric attributes contains all the linguistic information for the NPs in question.

2. A numerical value is assigned to each attribute according to its importance for establishing the anaphoric relation.

3. The list of numerical values assigned to each attribute constitutes the vector of the corresponding discourse entity.

The scalar product:

$$P_1.P_2 = \sum_{k}^{n} P_{1_k} \text{ x } P_{2_k}$$

(P_1 and P_2 are vectors with n elements, P_{1_k} is the k-element of P_1 and P_{2_k} - the k-element of P_2)

is used as a basic operation to compare the vectors of the discourse entities with that of the anaphoric expression. The result is a list of entities ordered according to the closeness of their vector to the vector of the anaphoric expression.

3.1.5 R. Mitkov: combination of linguistic and statistical methods

R. Mitkov describes an integrated model for anaphor resolution (Mitkov 1994a; Mitkov 1996a). The main idea incorporated into his anaphor resolution strategy is the combination of traditional linguistic methods with a new statistical approach for center tracking into a uniform architecture. Another important feature of the model is its restriction to a sublanguage/genre rather than concentrating on general natural language.

The model integrates modules containing different types of knowledge - syntactic, semantic, domain, discourse and heuristical. The syntactic module, for example, knows that the anaphor and antecedent must agree in number, gender and person. It checks to see if the c-command constraints hold and establishes disjoint reference. In cases of syntactic parallelism, it selects the noun phrase with the same syntactic function as the anaphor as the most probable antecedent.

The semantic module checks for semantic consistency between the anaphor and the possible antecedent. It filters out semantically incompatible candidates following the current verb semantics or the animacy of the candidate and gives preference to candidates with the same semantic role as the anaphor. The domain knowledge module is a knowledge base of the concepts of the domain considered and the discourse knowledge module can track the center of the current discourse segment (using a statistical Bayesian engine to suggest the most probable center on the basis of a new piece of evidence).

The syntactic and semantic modules usually filter the possible candidates and do not propose an antecedent (with the exception of syntactic and semantic parallelism). The proposal for an antecedent comes from the domain, heuristical, and discourse modules. The latter plays an important role in tracking the center and proposes it in many cases as the most probable candidate for an antecedent .

The program was tested in two modes (i) with the syntactic, semantic and domain modules activated and (ii) with the syntactic, semantic, domain *and* discourse modules activated. The results show an improvement in resolving anaphors when traditional linguistic approaches (syntactic and semantic constraints) are combined with the proposed statistical approach for tracking center (e.g. 89.1% vs. 87.7% , in the second case 86.7 vs. 91.6 accuracy).

3.1.6 Lappin and Leass' syntax-based approach

Shalom Lappin and Herbert Leass report (Lappin & Leass 1994) an algorithm for identifying the noun phrase antecedents of third person pronouns and lexical anaphors (reflexives and reciprocals). The algorithm (hereafter referred to as RAP (Resolution of Anaphora Procedure) applies to the syntactic representations generated by McCord's Slot Grammar parser (McCord 1990, 1993) and relies on salience measures derived from syntactic structure and a simple dynamic model of attentional state to select the antecedent noun phrase of a pronoun from a list of candidates. It does not employ semantic conditions (beyond those implicit in grammatical number and gender agreement) or real-world knowledge in choosing among the candidates.

RAP contains the following main components:

- An intrasentential syntactic filter for ruling out anaphoric dependence of a pronoun on an NP on syntactic grounds.
- A morphological filter for ruling out anaphoric dependence of a pronoun on an NP due to non-agreement of person, number, or gender features.
- A procedure for identifying pleonastic (non-anaphoric, semantically empty) pronouns.
- An anaphor binding algorithm for identifying the possible antecedent binder of a lexical anaphor within the same sentence.
- A procedure for assigning values to several salience parameters (grammatical role, parallelism of grammatical roles, frequency of mention, proximity, and sentence recency) for an NP. Higher salience weights are assigned to (i) subject over non-subject NPs, (ii) direct objects over other complements, (iii) arguments of a verb over adjuncts and objects of prepositional phrase adjuncts of the verb, and (iv) head nouns over complements of head nouns.
- A procedure for identifying anaphorically linked NPs as an equivalence class for which a global salience value is computed as the sum of the salient values of its elements.
- A decision procedure for selecting the preferred element of a list of antecedent candidates for a pronoun.

The authors tested the algorithm extensively on computer manual texts and conducted a blind test on a manual text containing 360 pronoun occurrences. The algorithm successfully identified the antecedent in 86% of the cases.

3.2 Alternative approaches

3.2.1 Nasukawa's "knowledge-independent" approach

Nasukawa's approach to pronoun resolution is relatively independent of external knowledge, making it highly practical. His approach is based on "preference according to existence of identical collocation patterns in the text", "preference according to the frequency of repetition in preceding sentences" as well as "preference according to syntactic position".

Nasukawa suggests that collocation patterns work as selectional constraints to determine how eligible a candidate for antecedent is. He makes use of a synonym dictionary and regards the synonym of an acceptable candidate as eligible itself.

Nasukawa finds, similarly to Mitkov (Mitkov 1993), that the frequency in preceding sentences of a noun phrase with the same lemma as the candidate may be an indication for preference when selecting an antecedent. Moreover, he considers a heuristic rule favouring subjects over objects (compare Mitkov 1993 where this preference is treated more exactly within a sublanguage). However, in his final implementation Nasukawa takes into account only syntactic (positional) preferences such as "candidate in a closer sentence" and "nearest candidate to the beginning of the same sentence" (preference over subject would have required prior syntactic analysis).

Each of the collocational, frequency or syntactic preferences gives its "preference value"; these values are eventually summed up. The candidate with the highest value is picked out as the antecedent.

As an evaluation corpus Nasukawa uses 1904 consecutive sentences (containing altogether 112 third-person pronouns) from eight chapters of two different computer manuals. His algorithm handles the pronoun "it" and has been reported to choose a correct antecedent in 93.8% of the cases.

3.2.2 Statistical/corpus processing approach (Dagan & Itai)

I. Dagan and A. Itai (Dagan & Itai 1990) report on a corpus-based approach for disambiguating pronouns which is an alternative solution to the expensive implementation of full-scale selectional constraints knowledge. They perform an experiment to resolve references of the pronoun "it" in sentences randomly selected from the corpus.

The model uses as selectional patterns the co-occurence patterns observed in the corpus. Candidates for antecedents are substituted for the anaphor and only those candidates available in frequent co-occurence patterns are approved of. The authors illustrate their approach on a sentence taken from the Hansard corpus:

> They knew full well that the companies held tax money$_k$ aside for collection later on the basis that the government$_i$ said it$_i$ was going to collect it$_k$.

There are two occurrences of "it" in the above sentence. The first is the subject of "collect" and the second is its object. Statistics are gathered for the three candidates for antecedents in the sentence: "money", "collection" and "government". The following table lists the patterns produced by substituting each candidate with the anaphor, and the number of times each of these patterns occurs in the corpus:

subject-verb	collection	collect	0
subject-verb	money	collect	5
subject-verb	government	collect	198
verb-object	collect	collection	0
verb-object	collect	money	149
verb-object	collect	government	0

According to the above statistics, "government" is preferred as the antecedent of the first "it" and "money" of the second.

This example demonstrates the case of definite semantic constraints which eliminate all but the correct alternative. According to the authors, several alternatives may satisfy the selectional constraints, and may be observed in the corpus a significant number of times. In such a case, the preference among several candidates should be performed by other means (which, however, are not discussed).

Dagan and Itai report an accuracy of 87% for the sentences with genuine "it" anaphors (sentences in which "it" is not an anaphor have been manually eliminated). It should also be pointed out that the success of this experiment depends on the parsing strategy implemented, which in this case was the K. Jennsen's PEG parser.

3.2.3 Connolly et al.'s machine learning approach (Connolly, Burger and Day 1994)

Connolly, Burger and Day (Connolly, Burger & Day 1994) describe a machine learning approach to anaphoric reference. Their idea is to cast anaphoric reference as a classification problem for which a classifier can be discovered empirically using traditional learning methods.

In order to apply machine learning techniques to the selection of the antecedent, the authors claim that the best defined problem which is suited to learning algorithms is the classification problem. The approach adopted in their research is to decompose the candidate selection problem into separate two-class classification problems. Each classification problem is defined on a pair of candidates and an anaphor, where the classes correspond to choosing one of the candidates as a "better" antecedent of the anaphor. By applying this classifier to successive pairs of candidates, each time retaining the best candidate, they effectively sort the candidates, thus choosing the best overall.

The classification itself is performed on a discrete-valued feature space. Instances are represented as attribute-value vectors, where attributes describe properties of the anaphor and the two candidates, and relationships among the three.

Once a classifier has been constructed, it is used by the reference resolution algorithm to select the best candidate using the following strategy. An initial instance is constructed by taking the anaphor and the first two candidates. This instance is given to the classifier, which indicates which candidate is the better one. The "losing" candidate is discarded, and another instance is constructed by pairing the "winner" with the next candidate. The new instance is now given to the classifier, which selects a new winner (which could be the same as the previous winner). This process continues until every candidate has been examined and the last winner is chosen as antecedent.

3.2.4 Aone & Bennett's machine learning approach (Aone & Bennet 1996)

C. Aone and S. Bennet describe an approach to building an automatically trainable anaphora resolution system (Aone & Bennet 1996). They tag anaphoric links in corpora of Japanese newspaper articles and use them as training examples for a machine learning algorithm (Quinlan's C4.5 decision tree algorithm).

The authors train a decision tree using feature vectors for pairs of an anaphor and its possible antecedent. Aone and Bennet use 66 training features such as lexical (e.g.

category), syntactic (e.g. grammatical role), semantic (e.g. semantic class) and positional (e.g. the distance between anaphor and antecedent) features. Those features can be either unary (i.e. features of either an anaphor or an antecedent such as syntactic number values) or binary (i.e. features concerning relations between the pairs such as the positional relation between an anaphor and an antecedent). Aone and Bennet employ different training methods using three parameters: anaphoric chains, anaphoric type identification and confidence factors.

The authors build 6 machine-learning based anaphora resolvers on the basis of training decision trees with 1971 anaphors. The evaluation is based on 1359 anaphors and two of the machine-learning resolvers demonstrate a precision close to 90%.

3.2.5 An uncertainty-reasoning approach (Mitkov 1995b)

Mitkov presents an Artificial Intelligence approach based on uncertainty reasoning. The rationale for the selection of such an approach is the following:

- in Natural Language Understanding, the program is likely to estimate the antecedent of an anaphor on the basis of incomplete information: even if information about constraints and preferences is available, it is natural to assume that a Natural Language Understanding program is not able to understand the input completely;

- since the initial constraint and preference scores are determined by humans, they are originally subjective and therefore should be regarded as uncertain facts;

The main idea is that the search for an antecedent can be regarded as an affirmation (or rejection) of the hypothesis that a certain noun phrase is the correct antecedent. Evidence is provided by "anaphor resolution symptoms" which have been empirically determined in previous research (Mitkov 1993). The availability/non-availability of each anaphor resolution symptom causes recalculation of the global hypothesis certainty factor CF_{hyp} (increase or decrease) until

$CF_{hyp} > CF_{threshold}$ for affirmation or
$CF_{hyp} < CF_{min}$ for rejection the hypothesis.

3.2.6 Two-engine approach (Mitkov 1997a)

Mitkov's two-engine approach is based on the interactivity of two engines which, separately, have been successful in anaphora resolution. The first engine incorporates the constraints and preferences of an integrated approach for anaphor resolution (Mitkov 1994a), while the second engine follows the principles of the uncertainty reasoning approach described in (Mitkov 1995b). The combination of a traditional and an alternative approach aims at providing maximum efficiency in the search for the antecedent.

The two-engine strategy evaluates each candidate for anaphor from the point of view of both the integrated approach and the uncertainty reasoning approach. If opinions coincide, the evaluating process is stopped earlier than would be the case if only one engine were acting. This also makes the searching process shorter: the preliminary tests show that the integrated approach engine needs about 90% of the search it would make when operating on its own; similarly, the uncertainty reasoning engine does only 67% of the search it would do when operating as a separate system. In addition, the results of using both approaches are more accurate, providing accuracy in some samples even higher than 90%.

3.2.7 Situational semantics approach (Tin & Akman 1994)

A situation semantics approach for anaphor resolution has been proposed by (Tin & Akman 1994) in which pronominal anaphors are resolved in a situation-theoretic computational environment by means of inference rules which operate on / unify with utterance situations

3.2.8 Using punctuation (Say & Vakman 1996)

Bilge Say and Vakman (Say & Vakman 1996) use punctuation within a DRT-framework as an additional constraint for anaphor resolution. They illustrate their approach with the example

> Jane$_i$, and Joe and Sue write books on England. If her$_i$ books are best-sellers then they are jealous.

3.3 Latest trends: knowledge-poor anaphora resolution

Most of the approaches outlined in sections 3.1 and 3.2 rely heavily on linguistic (and domain) knowledge. One of the disadvantages of developing a knowledge-based system, however, is that it is a very labour-intensive and time-consuming task. Consequently, the need for inexpensive and robust systems, possibly suitable for unrestricted texts, fuelled renewed research efforts (Baldwin 1997; Ferrandez et al. 1997; Kennedy & Boguraev 1996; Mitkov 1996b; Mitkov 1998b; Williams et al. 1996) in the field and a clear trend towards corpus-based and knowledge-poor approaches was established.

3.3.1 Kennedy and Boguraev's approach without a parser.

Kennedy and Boguraev’s (Kennedy & Boguraev 1996) approach is a modified and extended version of that developed by Lappin and Leass (Lappin & Leass 1994). Kennedy and Boguraev's system does not require "in-depth, full" syntactic parsing but works from the output of a part of speech tagger, enriched only with annotations of grammatical function of lexical items in the input text stream.

The basic logic of their algorithm parallels that of Lappin and Leass's algorithm. The determination of disjoint reference, however, represents a significant point of divergence between Kennedy and Boguraev's and Lappin and Leass's algorithms. The latter relies on syntactic configurational information, whereas the former, in the absence of such information, relies on inferences from grammatical function and precedence to determine the disjoint reference.

After the morphological and syntactic filters have been applied, the set of discourse referents that remain constitute the set of candidate antecedents for the pronoun. The candidate set is subjected to a final evaluation procedure which performs two functions: it decreases the salience of candidates which the pronoun precedes (cataphora is penalised), and increases the salience of candidates which satisfy either a locality or a parallelism condition, both of which apply to intrasentential candidates. The candidate with highest salience weight is determined to be the actual antecedent; in the event of a tie, the closest candidate is chosen. The approach works for both lexical anaphors (reflexives and reciprocals) and pronouns.

Evaluation reports 75% accuracy but this has to be given a "bonus" for these results span a very wide coverage: the evaluation was based on a random selection of genres, including press releases, product announcement, news stories, magazine articles, and other documents existing as World Wide Web pages.

3.3.2 Robust, knowledge-poor approach (Mitkov 1996, 1998).

Mitkov's robust approach works as follows: it takes as an input the output of a text processed by a part-of-speech tagger, identifies the noun phrases which precede the anaphor within a distance of 2 sentences, checks them for gender and number agreement with the anaphor and then applies the so-called antecedent indicators to the remaining candidates by assigning a positive or negative score (see below). The noun phrase with the highest aggregate score is proposed as antecedent; in the rare event of a tie, priority is given to the candidate with the higher score for immediate reference. If immediate reference has not been identified, then priority is given to the candidate with the best collocation pattern score. If this does not help, the candidate with the higher score for indicating verbs is preferred. If there is still no choice, the most recent from the remaining candidates is selected as the antecedent (for more details, see Mitkov 1998b).

The core of the approach lies in activating the antecedent indicators which play a decisive role in tracking down the antecedent from a set of possible candidates. Candidates are assigned a score (2,1,0,-1) for each indicator; the candidate with the highest aggregate score is proposed as the antecedent. The antecedent indicators have been identified on the basis of empirical studies and can be related to salience (definiteness/indefiniteness, givenness, indicating verbs, lexical reiteration, section heading preference, "non-prepositional" noun phrases, relative pronoun), to structural matches (collocation, immediate reference, sequential instructions), to referential distance or to preference of terms. Also, the indicators can be "impeding" (non-PP NPs, definiteness/indefiniteness), assigning negative scores to candidates or "boosting" (the rest), assigning positive scores. Whilst some of the indicators are more genre-specific (term preference) and others are less genre-specific ("immediate reference", "sequential instructions"), the majority appear to be genre-independent (Mitkov 1998; Mitkov et al. 1998). For instance, "definiteness/indefiniteness" considers definite noun phrases preceding the anaphor better candidates for antecedent than indefinite ones and therefore, indefinite noun phrases are penalised by the negative score of -1. Noun phrases in previous sentences and clauses representing the "given information" (theme) are deemed good candidates for antecedents and are assigned a score of 1. Also, noun phrases representing terms in the field are thought to be more likely antecedents than non-terms and are awarded a score of 1. Certain verbs (e.g. discuss, present, illustrate, identify etc.) are considered as quite indicative of the salience of the following noun phrase and therefore, the latter is given a score of 1. Lexical reiteration gives preference to lexically reiterated candidates (score 2 if repeated more than once and 1 otherwise), whereas section heading preference favours candidates which occur in the heading of a section (score 1). Collocation patterns give strong preference (score 2) to candidates which are found to appear more frequently (than other candidates) in the same collocational pattern with preceding or following verbs as the anaphor. The "immediate reference" indicator suggests that in constructions of the form "...(You) V_1 NP ... con (you) V_2 it (con (you) V_3 it)", where con {and/or/before/after...}, the noun phrase immediately after V_1 is a very likely candidate for an antecedent of the pronoun "it" immediately following V_2 and is therefore given preference (score 2 if the indicator holds and 0 otherwise) (As an illustration to this indicator, consider the example: To print the paper, you can stand the printer$_i$ up or lay it$_i$ flat). The indicator "non-prepositional noun phrases" operates on the evidence that noun phrases which are part of a prepositional phrase, are less likely candidates for antecedents and penalises them with the negative score of -1. Finally, on the basis of empirical observation, "referential distance" considers that in complex sentences, noun phrases in the previous clause[8] are the best candidates for the antecedent of an anaphor in the subsequent clause, followed by noun phrases in the previous sentence and by nouns situated 2 sentences prior to the sentence containing the anaphor (2, 1, 0). For anaphors in simple sentences, noun phrases in the previous sentence are the best candidate for antecedent, followed by noun phrases situated 2 sentences further back (1, 0).

The approach was evaluated against a corpus of technical manuals (223 pronouns) and achieved a success rate of 89.7%

3.1.4 Breck Baldwin's COGNIAC

CogNIAC is a system developed at the University of Pennsylvania to resolve pronouns with limited knowledge and linguistic resources (Baldwin 1997). The main assumption of CogNIAC is that there is a subclass of anaphora that does not require general purpose reasoning. The system requires for pre-processing its input sentence detection, part-of-speech tagging, simple noun phrase recognition, basic semantic category information.

CogNIAC is built on the following core rules:

1) Unique in discourse: if there is a single possible antecedent i in the read-in portion of the entire discourse, then pick i as the antecedent
2) Reflexive: pick the nearest possible antecedent in the read-in portion of current sentence if the anaphora is a reflexive pronoun
3) Unique in current and prior: if there is a single possible antecedent i in the prior sentence and the read-in portion of the current sentence, then pick i as the antecedent
4) Possessive pronoun: if the anaphor is a possessive pronoun and there is a single exact string match i of the possessive in the prior sentence, then pick i as the antecedent
5) Unique current sentence: if there is a single possible antecedent i the read-in portion of the current sentence, then pick i as the antecedent
6) If the subject of the prior sentence contains a single possible antecedent i, and the anaphor is the subject of the current sentence, then pick i as the antecedent

[8] The identification of clauses in complex sentences is done heuristically

CogNIAC operates as follows. Pronouns are resolved from left to right in the text. For each pronoun, the rules are applied in the presented order. For a given rule, if an antecedent is found, then the appropriate annotations are made to the text and no more rules are tried for that pronoun, otherwise the next rule is tried. If no rules resolve the pronoun, then it is left unresolved.

B. Baldwin reports 92% precision and 64% recall of his approach (Baldwin 1997). Its "resolve all" version which includes the lower precision rules (added to rules 1-6)

7) Backward center (Cb) -picking: If there is a Cb in the current finite clause that is also a candidate antecedent, then pick i as antecedent[9]
8) Pick the most recent: pick the most recent potential antecedent in the text

shows precision 77.9% on the training data (as opposed to Jerry Hobb's algorithm's 78.8%). The texts used for evaluation of CogNIAC were pre-processed by a part-of-speech tagger with noun phrases and finite clauses identified automatically as well. However, the pre-processing was manually corrected in order to be compatible with the results reported in the evaluation of Jerry Hobb's naive algorithm (Hobbs 1978).

4. Applications

4.1. Anaphora resolution in Machine Translation

4.1.1 Additional problems

The identification of the antecedents of anaphors is of crucial importance for correct translation. When translating into languages which mark the gender of pronouns, for example, it is essential to resolve the anaphoric relation. On the other hand, anaphor resolution is vital when translating discourse rather than isolated sentences since the anaphoric references to preceding discourse entities have to be identified. Unfortunately, the majority of Machine Translation systems do not deal with anaphor resolution and their successful operation usually does not go beyond the sentence level.

It is worthwhile mentioning that Machine Translation (MT) adds a further dimension to anaphor resolution. Additional complexity is due to gender discrepancies across languages, to number discrepancies of words denoting the same concept, to discrepancies in gender inheritance of possessive pronouns and discrepancies in target language anaphor selection (Mitkov & Schmidt 1998). The latter can be seen by the fact that even though in most of the cases the pronoun in the source language is translated by a target language pronoun (which corresponds in gender and number to the translation equivalent of the antecedent of the source language pronoun), there are some languages in which the pronoun is often translated directly by its antecedent (Malay). Additionally, pronominal anaphors are often elliptically omitted in the target language (Spanish, Italian, Japanese, Korean). Another interesting example is English-to-Korean translation. The English pronouns can be omitted elliptically, translated by a definite noun phrase, by their antecedent, or by one or two possible Korean pronouns, depending on the syntactic information and semantic class of the noun to which the anaphor refers (Mitkov et al. 1994, Mitkov et al. 1997).

[9]The Cb of an utterance is the highest ranked NP from the prior finite clause realised anaphorically in the current finite clause. The ranking used here is Subject > All other NPs

4.1.2 Research and development to date

Due to the fact that the majority of MT handle one-(simple)-sentence input, there is not an extensive amount of work reported to deal with anaphor resolution in MT.

4.1.2.1 *English-to-Japanese MT program (Wada 1990)*

Wada (Wada 1990) reports on an implementation of the Discourse representation theory in an LFG-based English-to-Japanese MT program. His anaphor resolution mechanism consists of three functional units:

- construction of the DRS
- storage of the salient element
- search for the antecedent

The first module constructs the DRSs compositionally, while the second module stores the most salient, focused element in the current discourse for processing of the next sentence. In order to find the most salient NP, this module sets three kinds of filters (grammatical function, use of pronominal and syntactic construction) and checks all the NPs appearing in the current sentence with respect to the three filters.

The third module incorporates three functions. The first function is "search", which searches for an antecedent by testing the accessibility on the DRS and syntactic constraints such as gender and number as well as binding features. If the search in the DRS fails, a further search in the module storing the salient element" is performed. According to the result of "search", three classes of pronominal anaphors are distinguished: (1) an antecedent is found in the current DRS (2) an antecedent is not found in the current DRS, but is controlled by a discourse focus and (3) an antecedent is not found either in the DRS or as the salient element.

The second function sets an unique anaphoric index in case (1), whereas the third function assigns 0 to the pronominal in (2) and undertakes a default word-for-word translation in (3).

4.1.2.2 English-to-Chinese Machine Translation (Chen 1992)

Chen (Chen 1992) describes the interpretation of overt and zero anaphors in English-to-Chinese and Chinese-to-English Machine Translation and justifies the importance of anaphor resolution. He outlines a "reflexive resolution algorithm" (based on c-command constraints and some semantic features), a "pronoun resolution algorithm" (based on c-command constraints and some simple semantic features) for overt anaphors in English and proposes an algorithm for the use of zero anaphors in Chinese. In addition, with a view to applying the results in Machine Translation transfer, he investigates the statistical distribution of anaphors (zero and pronominal) and their antecedents in both languages.

4.1.2.3 Resolution of Japanese zero pronouns (Nakaiwa et al. 1991, 1994, 1995)

Nakaiwa reports in various papers (Nakaiwa et al. 1994; Nakaiwa et al. 1995; Nakaiwa & Ikehara 1992; Nakaiwa & Ikehara 1995) the resolution of Japanese zero pronouns in Japanese-to-English Machine Translation. He uses semantic and pragmatic constraints such as conjunctions, verbal semantic attributes and modal expressions to determine intrasentential antecedents of Japanese zero anaphors with success in 98% of the cases (Nakaiwa & Ikehara 1995). Nakaiwa et al. also use additional information about text structure to resolve zero pronouns (Nakaiwa et al. 1996) and an evaluation based on 480 zero pronouns suggests a success rate of 93%.

4.1.2.4 Portuguese-to-English MT (Saggion & Carvalho 1994)

H. Saggion and A. Carvalho (Saggion & Carvalho 1994) describe pronoun resolution in a Portuguese-to-English MT system which translates scientific abstracts. They use syntactic agreement and c-command rules (Reinhard 1983) to solve intrasentential anaphors and syntactic analysis of the immediately preceding sentence plus a history list of previous antecedents (Allen 1987) to solve intersentential anaphors.

4.1.2.5 English-to-German MT system (Preuß et al. 1994)

Preuß, Schmitz, Hauenschild and Umbach (Hauenschild et al. 1993), (Preuß, Schmitz, Hauenschild & Umbach 1994) describe work on anaphor resolution in the English-to-German MT system KIT-FAST. Their approach uses two levels of text representation - structural and referential, as well as various "anaphor resolution factors" - proximity, binding, themehood, parallelism and conceptual consistency.

The structural text representation includes information about functor-argument relations (e.g. between nouns, verbs and adjectives and their complements), semantic roles of arguments (agent, affected, associated, location, aim), thematic structure of sentences, semantic features that express local or temporal conceptualisation, grammar and anaphoric relations represented by coindexation.

The referential text representation contains aspects of the text content, namely the discourse referents and the conceptual relations between them. Coreference relations are represented by one discourse referent and every relation that holds for an antecedent is also valid for an anaphor that refers to it. The referential information is represented in a terminological logic with the help of the knowledge representation system BACK (Weisweber 1994).

As far as the factors for anaphor resolution are concerned, proximity accounts for the fact that personal pronouns are most likely to have their antecedents in the superordinate or preceding sentence, while possessive pronouns tend to refer to a noun occurring in the same sentence. Binding excludes all sisters of a pronominal argument in the structural text representation as antecedent, whereas themehood defines structurally prominent constituents such as the subject or the topic of a sentence. Since the factors refer to the structural and referential representations, they have no access to purely syntactic information such as the subject. For this reason a notion of semantic subject is defined on the basis of the structural text representation and given preference. Moreover, the most topical candidate is preferred, whereas free adjuncts are considered as less likely antecedent candidates for possessive and personal pronouns.

The parallelism factor is expressed in agreement and identity of roles, whereas the conceptual consistency factor checks for compatibility between antecedents and anaphors.

4.1.2.6 English-to-Korean MT system (Mitkov et al. 1994)

Mitkov, Kim, Lee and Choi (Mitkov et al. 1994) describe an extension of the English-to-Korean MT MATES system with a view to tackling pronoun resolution. MATES is a transfer-based system, which does an English sentence analysis, transforms the result (parse tree) into an intermediate representation, and then transforms it into a Korean syntactic structure to construct a Korean sentence.

The paper argues that pronoun resolution cannot be ignored in English-to-Korean (however deceptive it may be in some cases) if aiming at good quality translation. The work suggests a number of practical rules for English-to-Korean anaphor transfer (including different types of complex sentences, general quantifiers, human 'it', non-human 'she', inanimate 'they') and outlines an anaphor resolution model for the English-to-Korean MT system MATES.

The anaphor resolution model proposed is a simplified version of the model described in (Mitkov 1994a). Full implementation of the latter, including center tracking inference engine, seems too costly for the immediate goals of the English-to-Korean MT system in operation.

4.1.2.7 Extension of CAT2 (Mitkov et al. 1995)

R. Mitkov, S.K.Choi and R. Sharp (Mitkov et al. 1995) describe an extension of the unification-based MT system CAT2. The latter was developed at IAI, Saarbrücken, as a sideline implementation to the Eurotra Project. The translation strategy is based on tree-to-tree transduction, where an initial syntactico-semantic tree is parsed, then transduced to an abstract representation ("interface structure"), designed for simple transfer to a target language interface structure. This is then transduced to a syntactico-semantic tree in the target language, whose yield provides the actual translated text.

The current version of the anaphora resolution model implemented in CAT2 is based exclusively on syntactic and semantic constraints and preferences. Syntax constraints include the almost obligatory agreement of the pronoun and its antecedent in number, person and gender as well as c-command constraints as formulated in (Ingria & Stallard 1989). Syntactic preferences used are syntactic parallelism and topicalisation. Syntactic parallelism gives preference to antecedents with the same syntactic role as the pronoun, whereas topicalised structures are searched first for possible anaphoric referents.

Semantic constraints and preferences include verb case role constraints, semantic networks constraints (semantic consistency) and semantic parallelism preference. Verb case role constraints stipulate that if filled by the anaphor, the verb case role constraints must also be satisfied also by the antecedent of the anaphor. Semantic networks indicate the possible links between concepts as well as concepts and their attributes; the semantic parallelism preference favours antecedents which have the same semantic role as the pronoun.

The project concentrates primarily on intersentential anaphor resolution, but since CAT2, like most other MT systems, only operates on single sentences, we simulate the intersententiality by conjoining sentences, as in:

The decision was adopted by the council; it published it.

The task is to resolve the pronominal references in the second "sentence" with the antecedents in the first. The implementation, which successfully handles pronominal resolution in the context of English-to-German translation, can be carried over to multiple-sentence input.

The noun phrase features relevant for anaphor resolution are collected in the complex feature "anaph", consisting of two additional features, "ref" (referential features) and "type". The referential features include the noun's agreement features (person, number, gender), lexical semantic features (e.g. animacy), and its referential index; the type feature indicates whether the noun is a pronoun or not:

anaph={ref={agr=A,sem=S, index=I},type=T}

All non-pronominal nouns receive a unique index by a special index generator within CAT2; each pronoun takes its index value by way of unification with its antecedent, as outlined below.

Anaphora resolution in CAT2 takes place in two main steps. First, all "anaph" features within a sentence are collected in a "cand" (candidates) feature and percolated to the S node, so that anaphora resolution between sentences can take place locally under the topmost node (Intrasentential anaphor resolution will already have occurred). Then, for each pronoun in the second sentence, its "ref" feature is resolved with those of the antecedents in the first sentence. Backtracking provides for all combinations, under the condition that the "ref" features agree, i.e. unify.

Illustrations of correct pronoun resolution are the translations by CAT2 of the sentences "The council adopted the decisions; the commission published them" (Der Rat verabschiedete die Beschlüsse; die Kommission veroeffentlichte sie.), "The council adopted the law; it published it." (Der Rat verabschiedete das Gesetz; er veröffentlichte es.), "The commission published the law; it was adopted by the council." (Die Kommission veröffentlichte das Gesetz; es wurde von dem Rat verabschiedet.) and the sentence "The decision was adopted by the council; it published it." (Der Beschluß wurde von dem Rat verabschiedet; er veröffentlichte ihn.)

4.2 Information extraction

Anaphora resolution in Information Extraction could be regarded as part of the more general task of coreference resolution which takes the form of merging partial data objects about the same entities, entity relationships, and events described at different discourse positions (Kameyama 1997).

Coreference resolution was introduced as a new domain-independent task at the 6th Message Understanding Conference (MUC-6) in 1995. This gave rise to a number of projects (Srivinas & Baldwin 1996, Gaizauskas et al. 1998, Kameyama 1997) dealing with coreference resolution in Information Extraction.

4.2.1 Kameyama's algorithm

Kameyama's algorithm was first implemented for the MUC-6 FASTUS system (Kameyama 1997) and produced one of the top scores (a recall of 59% and precision of 72%) in the MUC-6 Coreference Task. The input to the coreference resolver is incomplete syntactically (set of finite-state approximations of sentence parts) and poorer even than the input to Kennedy and Boguraev system (Kennedy & Boguraev 1996). The three main factors in Kameyama's algorithm are (i) accessible text regions, (ii) semantic consistency and (c) dynamic preference. The accessible text region for definite noun phrases is set to 10 sentences, and for pronouns - 3 sentences (ignoring paragraph boundaries and no antecedents beyond this limit are considered). The semantic consistency filters are number consistency, sort consistency (anaphors must either be of the same sort with their antecedents or subsume their sort; e.g. "company" subsumes "automaker" and "the company" can take "a Chicago-based automaker" as an antecedent) and modifier consistency (e.g. French and British are inconsistent but French and multinational are consistent). The basic underlying hypothesis of the dynamic preference is that intersentential candidates are more salient than intrasentetial ones and that fine-grained syntax-based salience fades with time. Since fine-grained syntax with grammatical functions is unavailable in the input to the resolver, the syntactic prominence of subjects, for instance, is approximated by left-right linear ordering.

4.2.2 Pronoun resolution in M-LaSIE (Azzam et al. 1998)

Gaizauskas et al.'s algorithm was developed for the M-LaSIE multilingual IE system (Gaizauskas et. al. 1997). Unlike many IE systems that skim texts and use large collections of shallow, domain-specific patterns and heuristics to fill in templates, M-LaSIE attempts a fuller analysis, first translating individual sentences to a quasi-logical form (QLF), and then constructing a weak discourse model of the entire text. After the QLF representation of a sentence has been added to the discourse model, all new indices (those produced by that sentence) are compared with other indices in the model to determine whether any pair has merged, representing a coreference in the text. New indices introduced by pronouns are compared with existing indices from the current paragraph, and then each previous paragraph in turn until an antecedent is found.

In addition to the base coreference algorithm, an experiment with an approach based on Azzam's proposed extensions (Azzam, 1996) to Sidner's focusing approach has been carried out. The evaluation of pronoun resolution based on a small English text sample (19 pronouns) suggests a recall of 63% and precision 86%.

4.3 Other applications

4.3.1 Automatic abstracting

Researchers in automatic abstracting are interested more and more in anaphora resolution since techniques for extracting important sentences are more accurate if anaphoric references of indicative concepts / noun phrases are taken into account as well. Badwin and Morton describe an abstracting approach based on coreference (Baldwin & Morton 1998): coreference chains play a decisive role in computing the score of each sentence and sentences with highest scores are selected as an abstract. Also, Boguraev and Kennedy propose an approach to content characterisation of text documents (Boguraev & Kennedy 1997) which is based on their anaphora resolution model (Kennedy & Boguraev 1996).

5. Some new developments and issues that need further attention

5.1 Evaluation in anaphora resolution

Against the background of growing interest in the field, it seems that insufficient attention has been paid to the evaluation of the systems developed. Even though the number of works reporting extensively on evaluation in anaphora resolution is increasing (Aone & Bennet 1996; Azzam et al. 1998; Baldwin 1997; Gaizauskas & Humphreys 1996; Lappin & Leass 1994, Mitkov & Stys 1997, Mitkov et al. 1998), the forms of evaluation that have been proposed are not sufficient or perspicuous.

As in any other NLP task, evaluation is of crucial importance to anaphora resolution. The MUC (Message Understanding Conference) initiatives suggested the measures "recall" and "precision" be used for evaluating the performance of coreference (but also for anaphora) resolution systems. These measures have proved to be useful indicators and they have already been used in some of the above mentioned works (Azzam et al. 1998; Baldwin 1997, Gaizauskas & Humphreys 1996).

In anaphora resolution, the recall would be defined as the ratio number of correctly resolved anaphors / number of all anaphors and precision - as the ratio number of correctly resolved anaphors / number of anaphors upon which resolution is attempted (see also Aone & Bennet 1996; Baldwin1997)[10]. Therefore, if the approach is robust and proposes an antecedent for each pronominal anaphor, the success rate (successfully resolved anaphors / number of all anaphors) would be equal to both recall and precision.

It is felt, however, that evaluation in anaphora resolution needs further attention. Measuring the success rate of an anaphora resolution system in terms of "recall" and "precision" is undoubtedly an important (and consistent) step in assessing the efficiency of anaphora resolution approaches, but as we have already pointed out, they cannot be seen as distinct measures for robust systems. In addition, it appears that they alone

[10]The above definitions refer to the resolution process only and not to the process of identification of anaphors which of course will have "recall" and "process" defined differently.

cannot provide a comprehensive overall assessment of an approach. In order to see how much a certain approach is "worth", it would be necessary to assess it against other "benchmarks", e.g. against other existing or baseline models. It also makes sense to evaluate the performance on anaphors which do not have sole candidates for antecedents and whose antecedents cannot be identified on the basis of gender and number agreement only (see the notion of "critical success rate", Mitkov 1998a).

In the absence of a wide range of "universal" benchmarks which could serve as evaluation samples for the different anaphora resolution approaches, Mitkov proposes (Mitkov 1998a) that evaluation would be more complete if the following additional evaluation tasks were carried out[11]:

- comparative evaluation of baseline models
- comparative evaluation of other similar methods
- comparative evaluation of other well-known methods
- evaluation of the performance on anaphors which cannot be identified on the basis of gender and number agreement only (critical success rate)

Furthermore, the evaluation would be more revealing if in addition to evaluating a specific approach as a whole, we broke down the evaluation process by looking at the different components involved. In the case of anaphora resolution, evaluation of each individual factor employed in the resolution process has been proposed in (Mitkov 1998a). Such an evaluation would provide important insights into how the overall performance of the system could be improved[12] (e.g. through changing the weights/scores of the factors). The same work proposes the notion of "discriminative power" of anaphora resolution factors which can play an important role in preferential architectures (for more details on factors, preferences and constraints, see Mitkov, 1997b).

5.2 Multilingual anaphora resolution

5.2.1 Anaphora resolution and languages covered

Most of anaphora resolution approaches so far developed have had a monolingual orientation (for recent multilingual developments see next subsection), concentrating on pronoun resolution in one language alone. As in other NLP areas, anaphora resolution approaches and implementations for English considerably outnumber those for other languages. Anaphora resolution projects have been reported for French (Popescu-Belis & Robba 1997, Rolbert 1989), German (Dunker & Umbach 1993; Fischer et al. 1996; Leass & Schwall 1991; Stuckardt 1996; Stuckardt 1997), Japanese (Aone & Bennet 1996; Mori et al. 1997; Nakaiwa & Ikehara 1992; Nakaiwa & Ikehara 1995; Nakaiwa et al. 1995; Nakaiwa et al. 1996; Wakao 1994), Portuguese (Abraços & Lopes 1994), Spanish (Ferranndez et al. 1997), Swedish (Fraurud, 1988) and Turkish (Tin & Akman, 1994) and it looks as if more languages will be covered soon (Mitkov et al. 1998).

5.2.2 Recent multilingual developments

---

[11] Mitkov describes the evaluation of his robust approach against these additional tasks in (Mitkov 1998a)

[12] This is in line with M. Walker's comments (Walker 1989) that evaluation contributes to (i) obtaining general results on a methodology for doing evaluation and (ii) discovering ways we can improve current theories (approaches).

The developments in anaphora resolution take place in the wider context of NLP where the search for multilingual applications is a live issue. Against the background of growing interest in multilingual work, it was surprising that until recently, no anaphora resolution project[13] had looked at the multilingual aspects of the approaches that have been developed, or, in particular, at how a specific approach could be used or adapted for other languages.

The last few months, however, have seen the emergence of the first multilingual anaphora resolution projects (Mitkov & Stys 1997; Mitkov et al. 1998; Azzam et al. 1998) and therefore, the establishment of a new trend towards multilinguality in the field.

In his recent works (Mitkov 1998c; Mitkov et al. 1998) R. Mitkov reports a project which has a truly multilingual character. His robust, knowledge-poor approach (Mitkov 1998a; see also subsection 3.3.2 of the current paper) was initially developed and tested for English, but was also adapted and tested for Polish and Arabic. It was found that the approach could be adapted with minimum modification for both languages and moreover, even if used without any modification, it still delivered acceptable success rates. Evaluation shows a success rate of 89.7% for English, 93.3% for Polish and 95.8% for Arabic. The approach is currently being tested for Finnish, French and Russian (Mitkov 1998c).

5.3 Anaphora resolution in the world of growing resources

Annotation of referential links has not yet benefited from the level of automation enjoyed by its lexical, syntactical and semantic "counterparts". For instance, part-of-speech tagging has shown remarkable accuracy (99.2% see Voutilainen 1995) and word sense disambiguation has reported accuracy of more than 96%[14] (Yarowsky 1995). Even though no practical domain-independent parser of unrestricted texts has yet been developed (Carroll & Briscoe 1996a), there are robust domain-independent parsers which are able to parse around 80% of the sentences from general text corpora with bracketing recall and precision in the region of 83% (Carroll & Briscoe 1996b); parsers specific to particular corpora or text genres (e.g. Collins 1997) can achieve 100% coverage and up to 87% recall and precision. However, "referential tagging" has not been fully explored (and developed) and this is due, no doubt, to the complexity of automatic anaphora (or more general, coreference) resolution.

One of the best known tools for annotation of anaphoric links in a corpus is XANADU - an X-windows interactive editor written by Roger Garside, which offers the user an easy-to-navigate environment for manually marking pairs of anaphors-antecedents (Fligelstone 1992). In spite of its attractive features, XANADU is still based on manual antecedent identification and labelling. Manual annotation, however, imposes a considerable demand on human time and labour.

R. Mitkov has put forward the idea of incorporating his robust, knowledge-poor approach to anaphora resolution (Mitkov 1997c) within a larger architecture for rough

---

[13]Except for Aone and McKee (Aone & McKee 1993); their paper, however, does not report evaluation of anaphora resolution in the languages mentioned (English, Japanese, Spanish)

[14]This accuracy, however, has been challenged in personal communication by other authors on the basis that it has been tested on a limited number of words only; evaluation results on larger samples still show accuracy of over 94% (Wilks & Stevenson 1998)

automatic annotation of anaphoric links in corpora. The proposal deals with pronominal anaphora only and "rough annotation" implies that a follow-up manual correction would be necessary.

Automatic referential links in corpora is a highly attractive research task that will definitely need further attention in the future.

5.4 A few open questions

There are a number of questions that remain unanswered or need further attention and which are directly related to the notion of factors in anaphora resolution such as (i) how dependent are factors? (ii) do factors hold good for all genres? (iii) how multilingual can factors be? and (vi) which is the best order to apply the factors?

5.4.1 Dependence and mutual dependence of factors

While it is vital to arrive at a comprehensive list of contributory factors (or a core set of maximally contributory factors), it is worthwhile to consider not only the impact of each factor on the resolution process but also the impact of factors on other factors. A key issue which needs further attention is the "(mutual) dependence" of factors.

In order to clarify the notion of (mutual) dependence, it would be helpful to view the "factors" as "symptoms", "indicators" i.e. as "present" or "absent" in a certain discourse situation. For instance, if "gender agreement" holds between a candidate for an anaphor and the anaphor itself, the symptom or indicator "gender agreement" is said to be present. Similarly, if the candidate is in a subject position, the symptom "subjecthood" is said to be present.

Dependence/mutual dependence of factors is defined in the following way. Given the factors x and y, the factor y is said to be dependent on factor x to the extent that the presence of x implies y. Two factors will be termed mutually dependent if each depends on the other.

The phenomenon of (mutual) dependence has not yet been fully investigated (see Mitkov 1997b), but it is felt that it can play an important role in the process of anaphora resolution, especially in algorithms based on the ranking of preferences. Information on the degree of dependence would be especially welcome in a comprehensive probabilistic model and will undoubtedly lead to more precise results.

5.4.2 Do the factors hold good for all genres?

It is believed that it is important to investigate whether or not a specific factor is genre-specific or if it holds good for all genres. Such information could prove to be very helpful in designing and developing a genre-specific anaphora resolution system. For instance, Mitkov's genre-specific robust pronoun resolution approach which is based on a number of genre-specific indicators, shows a high success rate in the range of 90% for the genre of technical manuals (Mitkov 1998b).

5.4.3 Factors and multilinguality

Do factors hold equally good for all languages? Preliminary findings (Mitkov et al. 1998) on the basis of a few languages (English, Arabic, Polish, Russian, Finnish) suggest a negative answer to this question but further study into other languages would be of interest and will have a direct implication for the development of multilingual anaphora resolution systems. There are certain factors that work well for all languages (e.g. collocation) and factors that are typically monolingual (e.g. the "relative pronoun indicator" which is good for Arabic only, see Mitkov et al. 1998). On the other hand, multilingual factors do not work equally well for all languages. For instance, the typically multilingual constraint of number agreement is more obligatory for some languages and less obligatory for others (Mitkov 1998b). Also, referential distance plays a more important role in Arabic than in English.

In a multilingual robust pronoun resolution system Mitkov and colleagues (Mitkov et al. 1998) used a core set of multilingual factors (whose scores may differ for each language, though) as a basic platform which is enhanced by additional language-specific factors each time a language (English, Arabic, Polish) is activated.

5.4.4 Order of constraints and priority of preferences

Does order of constraints matter? Since "absolute" constraints have to be met, not complying with any of them means discounting candidates. Therefore, it is apparent that the order in which the constraints are applied does not matter.

In a system which incorporates both constraints and preferences, it would be natural to start with constraints and then to switch to preferences. It is feared, however, that unless there is a comprehensive list of exceptions, simply discounting candidates on the basis of gender and number agreement in English could be risky (there are a number of cases where collective nouns may be referred to by plural pronouns and cases where plural nouns may be referred to by a singular pronoun[15]). Therefore, unless we have such a comprehensive list, our personal inclination would be to rely on a preferences-based architecture.

As far as preferences are concerned, it would be natural to start with the more "contributory" factors in terms of numerical value or perhaps even better, with the most discriminative factors. In our experiments so far we have tried both descending (starting first the higher value factors) and ascending orders of application. We did not find any essential difference in the final result. However, the searching process in the second option was, as expected, longer.

For preferences-based computational strategies which employ all preferences to reach a judgement as to whether a certain NP should be accepted as an antecedent or not, the order of application of factors does not matter at all since the final decision is made on the basis of the aggregate score generated by the scoring formula.

References

[15] In English there are certain collective nouns which do not agree in number with their antecedents e.g. "government", "team", "parliament" etc. can be referred to by "they"; equally some plural nouns such as "data" can be referred to by "it"

Abraços, Jose & José G. Lopes. 1994. "Extending DRT with a focusing mechanism for pronominal anaphora and ellipsis resolution". *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94)*, 1128-1132, Kyoto, Japan.

Allen, James. 1987. *Natural language understanding*. The Benjamin/Cummings Publishing Company Inc.

Aone, Chinatsu & Douglas McKee. 1993. "A language-independent anaphora resolution system for understanding multilingual texts". *Proceedings of the 31st Annual Meeting of the ACL (ACL'93)*, 156-163.

Aone, Chinatsu & Scot W. Bennet. 1996. "Applying machine learning to anaphora resolution". *Connectionist, statistical and symbolic approaches to learning for Natural Language Processing* ed. by S. Wermter, E. Riloff & G. Scheler, 302-314. Berlin: Springer.

Azzam, Saliha. 1996. "Resolving anaphors in embedded sentences". *Proceedings of the 34th Meeting of the Association fr Computational Linguistics (ACL'96)*, 263-269. Santa Cruz, USA.

Azzam, Saliha, Kevin Humphreys & Robert Gaizauskas. 1998. “Coreference resolution in a multilingual information extraction”. *Proceedings of the Workshop on Linguistic Coreference*. Granada, Spain.

Baldwin, Breck. 1997. "CogNIAC: high precision coreference with limited knowledge and linguistic resources". *Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution*, 38-45, Madrid, Spain.

Baldwin, Breck & Thomas S. Morton. 1998. “Dynamic coreference-based summarization”. *Proceedings of the Third International Conference on Empirical Methods in Natural Language Processing (EMNLP-3)*, 1-6. Granada, Spain.

Bobrow, D.G. 1964. *A question-answering system for high school algebra word problems*. Proceedings of AFIPS conference, 26, FJCC.

Boguraev, Branimir & Christopher Kennedy. 1997. “Salience-based content characterisation of documents”. *Proceedings of the ACL'97/EACL'97 workshop on Intelligent scalable text summarisation*, 3-9. Madrid, Spain

Brennan, S., M. Friedman and C. Pollard. 1987. A centering approach to pronouns. *Proceedings of the 25th Annual Meeting of the ACL (ACL'87)*, 155-162. Stanford, CA, USA.

Carbonell, James G. & Ralf D. Brown. 1988. "Anaphora resolution: a multi-strategy approach". *Proceedings of the 12. International Conference on Computational Linguistics (COLING'88)*, Vol.I, 96-101, Budapest, Hungary.

Carroll, John & Ted Briscoe. 1996a. "Robust Parsing - a brief overview". *Proceedings of the Workshop on Robust Parsing at the 8th European Summer School in Logic, Language and Information (ESSLLI'96)*, 1-7, Report CSRP 435, COGS. Brighton: University of Sussex.

Carroll, John & Ted Briscoe. 1996b. "Apportioning development effort in a probabilistic LR parsing system through evaluation". *Proceedings of the ACL SIGDAT Conference on Empirical Methods in Natural Language Processing*, 92-100, Philadelphia, USA.

Carter, David M. 1986. *A shallow processing approach to anaphor resolution*. PhD thesis. University of Cambridge.

Carter, David M. 1987. *Interpreting anaphora in natural language texts*. Chichester: Ellis Horwood

Chen, Hsin-Hsi. 1992. "The transfer of anaphors in translation". *Literary and Linguistic Computing, 7 (4)*, 231-238.

Collins, Michael 1997. "Three generative, lexicalised models for statistical parsing". *Proceedings of the 35th Annual Meeting of the ACL (ACL'97)*, 16-23, Madrid, Spain.

Connoly, Dennis, John D. Burger & David S. Day. 1994. "A Machine learning approach to anaphoric reference". *Proceedings of the International Conference "New Methods in Language Processing"*, 255-261, Manchester, United Kingdom.

Dagan, Ido & Alon Itai. 1990. "Automatic processing of large corpora for the resolution of anaphora references". *Proceedings of the 13th International Conference on Computational Linguistics (COLING'90)*, Vol. III, 1-3, Helsinki, Finland.

Dahl, D. & Catherine Ball. 1990. *Reference resolution in PUNDIT*. Research Report CAIT-SLS-9004. Paoli: Center for Advanced Information Technology.

Dunker, Guido & Carla Umbach. 1993. *Verfahren zur Anaphernresolution in KIT-FAST*. Internal Report KIT-28, Technical University of Berlin.

Ferrandez, Antonio, Manolo Palomar and Moreno L. 1997. "Slot unification grammar and anaphora resolution". *Proceedings of the International Conference on Recent Advances in Natural Language Proceeding (RANLP'97)*, 294-299. Tzigov Chark, Bulgaria.

Fischer, Ingrid, Bernd Geistert & Günter Goerz 1996. "Incremental anaphora resolution in a chart-based semantics construction framework using l-DRT". *Proceedings of the International Colloquium on Discourse Anaphora and Anaphora Resolution. Lancaster (DAARC)*, 235-244, Lancaster, UK.

Fligelstone, Steve. 1992. "Developing a scheme for annotating text to show anaphoric relations". *New Directions in English Language Corpora: Methodology, Results, Software Developments*. ed. by Gerhard Leitner, 153-170. Berlin: Mouton de Gruyter.

Fraurud, Kari. 1988. "Pronoun Resolution in unrestricted text". *Nordic Journal of Linguistics* 11, 47-68

Gaizauskas, Robert & Kevin Humphreys. 1996. Quantitative evaluation of coreference algorithms in an information extraction system. Paper presented at the *Discourse Anaphora and Anaphor Resolution Colloqium (DAARC*), Lancaster, UK

Gaizauskas, Robert, Kevin Humphreys, Saliha Azzam & Yorick Wilks. 1997. Conceptions vs. lexicons: an architecture for multilingual information extraction. In M.T. Pazienza, editor, *Proceedings of the Summer School on Information Extraction* (SCIE-97), 28-43. Springer-Verlag.

Grosz, Barbara J., Aravind Joshi & Scott Weinstein. 1983. "Providing a unified account of definite noun phrases in discourse". *Proceedings of the 21. Annual meeting of the Association for Computational Linguistics (ACL'83)*, 44-50.

Grosz, Barbara J., Aravind Joshi & Scott Weinstein. 1995. "Centering: a framework for modelling the local coherence of discourse". *Computational Linguistics*, 21(2), 44-50.

Haegeman, Liliane. 1994. *Introduction to Government and Binding* Theory. Oxford: Blackwell.

Halliday, M. & R. Hasan. 1996. *Cohesion in English*. (Longman English Language Series 9). London: Longman

Hirst, Graeme. 1981.*Anaphora in natural language understanding*. Berlin Springer Verlag, 1981.

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

Hobbs, Jerry R. 1978 "Resolving pronoun references". *Lingua*, 44, 339-352.

Huddleston, Rodney. 1984. *Introduction to English Grammar*. Cambridge: Cambridge University Press.

Ingria, Robert J.P. & David Stallard. 1989. "A computational mechanism for pronominal reference". *Proceedings of the 27th Annual Meeting of the ACL*, 262-271, Vancouver, British Columbia.

Kameyama, Megumi. 1997. "Recognizing referential links: an information extraction perspective" *Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution*, 46-53. Madrid, Spain.

Karlsson, Fred, Atro Voutilainen, Juha Heikkilä & Arto Anttila. 1995. *Constraint grammar: a language-independent system for parsing free text.* Berlin/New York: Mouton de Gruyter

Kennedy, Christopher & Branimir Boguraev, 1996. "Anaphora for everyone: pronominal anaphora resolution without a parser". *Proceedings of the 16th International Conference on Computational Linguistics (COLING'96)*, 113-118. Copenhagen, Denmark

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

Leass Herbert & Ulrike Schwall. 1991. *An anaphora resolution procedure for machine translation*. IBM Germany Science Center. Institute for Knowledge Based Systems, Report 172

McCord, Michael. 1990. "Slot grammar: a system for simpler construction of practical natural language grammars." *Natural Language and Logic: International Scientific Symposium*, edited by R. Studer, 118-145. Lecture Notes in Computer Science, Berlin: Springer Verlag.

McCord, Michael. 1993. "Heuristics for broad-coverage natural language parsing." *Proceedings of the APRA Human Language Technology Workshop*, University of Pennsylvania.

Mitkov, Ruslan. 1993. *Anaphora resolution and natural language understanding*. Unpublished internal report. Universiti Sains Malaysia, Penang.

Mitkov, Ruslan. 1994a. "An integrated model for anaphora resolution". *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94)*, 1170-1176, Kyoto, Japan.

Mitkov, Ruslan. 1994b. "A new approach for tracking center". Proceedings of the International Conference "New Methods in Language Processing" (NeMLaP-1). Manchester, UK.

Mitkov, Ruslan. 1995a. *Anaphora resolution in Natural Language Processing and Machine Translation.* Working paper. Saarbrücken: IAI

Mitkov, Ruslan. 1995b. "An uncertainty reasoning approach for anaphora resolution". *Proceedings of the Natural Language Processing Pacific Rim Symposium (NLPRS'95)*, 149-154, Seoul, Korea.

Mitkov, Ruslan. 1996a. "Anaphora resolution: a combination of linguistic and statistical approaches". *Proceedings of the Discourse Anaphora and Anaphor Resolution (DAARC'96)*. Lancaster, UK.

Mitkov, Ruslan. 1996b. "Pronoun resolution: the practical alternative". Paper presented at the *Discourse Anaphora and Anaphor Resolution Colloquium (DAARC*), Lancaster, UK.

Mitkov, Ruslan. 1997a. "Two engines are better than one: generating more power and confidence in the search for the antecedent" In R. Mitkov, N. Nicolov (Eds) "*Recent Advances in Natural Language Processing*". John Benjamin Publishers.

Mitkov, Ruslan. 1997b. "Factors in anaphora resolution: they are not the only things that matter. A case study based on two different approaches" *Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution.*, 14-21. Madrid, Spain.

Mitkov, Ruslan. 1997c. "How far are we from (semi-)automatic annotation of anaphoric links in corpora?" *Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution*, 82-87. Madrid, Spain.

Mitkov, Ruslan. 1998a. "Evaluating anaphora resolution approaches". *Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium (DAARC'2).* Lancaster, UK

Mitkov, Ruslan. 1998b. "Robust pronoun resolution with limited knowledge". *Proceedings of the 18.th International Conference on Computational Linguistics (COLING'98)/ACL'98 Conference*. Montreal, Canada.

Mitkov, Ruslan. 1998c. "The latest in anaphora resolution: going robust, knowledge-poor and multilingual". *Proceedings of the Annual Conference of the Spanish Association of Computational Linguistics (SEPLN'98)* (invited talk - forthcoming). Alicante, Spain.

Mitkov, Ruslan & Malgorzata Stys. 1997. "Robust reference resolution with limited knowledge: high precision genre-specific approach for English and Polish". *Proceedings of the International Conference "Recent Advances in Natural Language Proceeding" (RANLP'97)*, 74-81. Tzigov Chark, Bulgaria.

Mitkov, Ruslan & Paul Schmidt. 1998. "On the complexity of anaphora resolution in Machine Translation". *Mathematical and computational analysis of natural language* ed. by Carlos Martin-Víde. Amsterdam: John Benjamins

Mitkov Ruslan, Sung-Kwon Choi & Randy Sharp. 1995. "Anaphora resolution in Machine Translation". *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI'95)*, Leuven, Belgium.

Mitkov, Ruslan, Lamia Belguith & Malgorzata Stys. 1998. "Multilingual robust anaphora resolution". *Proceedings of the Third International Conference on Empirical Methods in Natural Language Processing (EMNLP-3),* 7-16. Granada, Spain.

Mitkov, Ruslan, Hiongun Kim, Kang-Hyuk Lee & Key-Sun Choi. 1994. "The lexical transfer of pronominal anaphors in Machine Translation: the English-to-Korean case". *Proceedings of the SEPLN'94 conference*, Cordoba, Spain.

Mitkov, Ruslan, Kang-Hyuk Lee, Hiongun Kim & Key-Sun Choi. 1997. “English-to-Korean Machine Translation and the problem of anaphora resolution”. *Journal of Literary and Linguistics Computing*, 12(1), 23-30.

Mori, Tatsunori, Mamoru Matsuo, Hiroshi Nakagawa. 1997. Constraints and defaults of zero pronouns in Japanese instruction manuals. *Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution*, 7-13. Madrid, Spain

Nakaiwa, Hiromi & Satoru Ikehara. 1992. "Zero pronoun resolution in a Japanese-to-English Machine Translation system by using verbal semantic attributes". *Proceedings of 3rd Conference on Applied Natural Language Processing (ANLP'92)*. Trento, Italy.

Nakaiwa, Hiromi & Satoru, Ikehara. 1995. "Intrasentential resolution of Japanese zero pronouns in a Machine Translation system using semantic and pragmatic constraints". *Proceeedings of the International Conference on Theoretical and Methodological Issues in Machine Translation (TMI'95)*, 96-105, Leuven, Belgium.

Nakaiwa, Hiromi, A. Yokoo, A. & Satoru Ikehara. 1994. "A system of verbal semantic attributes focused on the syntactic correspondence between Japanese and English". *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94)*. Kyoto, Japan.

Nakaiwa, Hiromi, S. Shirai, Satoru Ikehara & T. Kawaoka. 1995. "Extrasentential resolution of Japanese zero pronouns using semantic and pragmatic constraints". *Proceedings of the AAAI 1995 Spring Symposium Series: Empirical methods in discourse interpretation and generation*.

Nakaiwa, Hiromi, Francis Bond, Takahiro Uekado & Yayoi Nozawa. 1996. "Resolving zero pronouns in texts using textual strucure". *Proceedings of the NeMLaP-2 conference*. Ankara, Turkey.

Nasukawa, Tetsuya. 1994. "Robust method of pronoun resolution using full-text information". *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94)*, 1157-1163, Kyoto, Japan.

Popescu-Belis, Andrei & Isabelle Robba. 1997. "Cooperation between pronoun and reference resolution for unrestricted texts". *Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution*, 30-37. Madrid, Spain.

Preuß S., B. Schmitz, C. Hauenschild & C. Umbach . 1994. “Anaphora Resolution in Machine Translation”. *Studies in Machine Translation and Natural Language Processing*, (Volume 6 "Text and content in Machine Translation: Aspects of discourse representation and discourse processing") ed. by W. Ramm (ed), 29-52. Luxembourg: Office for Official Publications of the European Community.

Reinhardt, Tanya. 1983. *Anaphora and semantic interpretation.* Chicago III: University of Chicago Press.

Rich, Elaine & Susann LuperFoy. 1988. "An architecture for anaphora resolution". *Proceedings of the Second Conference on Applied Natural Language Processing (ANLP-2)*, 18-24, Texas, U.S.A.

Rico Pérez, Celia. 1994. "Resolución de la anáfora discursiva mediante una estrategia de inspiración vectoral". *Proceedings of the SEPLN'94 conference (SEPLN'94)*. Cordoba, Spain.

Rolbert, Monique. 1989. *Résolution de formes pronominales dans l'interface d'interogation d'une base de données*. Thèse de doctorat. Faculté des sciences de Luminy.

Saggion, Horacio & Ariadne Carvalho. 1994. “Anaphora resolution in a machine translation system”. *Proceedings of the International Conference "Machine translation, 10 years on"*, 4.1-4.14. Cranfield, UK.

Say, Bilke & Varol Akman. 1996. "An information-based treatment of punctuation". *International Conference on Mathematical Linguistics (IMCL'96)*, Tarragona, Spain.

Sidner, Candy L. 1979. *Towards a computational theory of definite anaphora comprehension in English discourse*. Technical Report No. 537. M.I.T., Artificial Intelligence Laboratory.

Sidner , Candy L. 1986 "Focusing in the comprehension of definite anaphora". *Readings in Natural Language Processing* ed. by B. Grosz, K. Jones & B. Webber. Morgan Kaufmann Publishers.

Srivinas Anoop & Breck Baldwin. 1996. “Exploiting super tag representation for fast coreference resolution”. *Proceedings of the NLP+ IA’96 conference*, 263-269. Moncton, Canada.

Stuckardt, Roland. 1996. "An interdependency-sensitive approach to anaphor resolution". *Proceedings of the International Colloquium on Discourse Anaphora and Anaphora Resolution. Lancaster (DAARC)*, 400-413. Lancaster, UK.

Stuckardt, Roland. 1997. "Resolving anaphoric references on deficient syntactic descriptions". *Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution*, 30-37. Madrid, Spain.

Stys, Malgorazata & Stefan Zemke. 1995. "Incorporating discourse aspects in English - Polish MT: towards robust implementation". *Proceedings of the international conference "Recent Advances in Natural Language Processing" (RANLP'95)*. Tzigov Chark, Bulgaria.

Tin, Erkan & Varol, Akman. 1994. "Situated processing of pronominal anaphora". *Proceedings of the KONVENS"94 Conference*, 369-378, Vienna, Austria.

Voutilainen, Atro. 1995. “A syntax-based part-of-speech tagger”. *Proceedings of the 7th conference of the European Chapter of the Association for Computational Linguistics (EACL’95)*, 157-164. Dublin, Ireland.

Voutilainen, Atro, Juha Heikkilä & Arto Anttila. 1992. *A constraint grammar of English: a performance-oriented approach*. Publication No. 21, Helsinki: University of Helsinki

Yarowsky, David. 1995. “Unsupervised word sense disambiguation rivaling supervised methods”. *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL '95)*, 189-196, Cambridge, MA.

Wada, Haime. 1990. “Discourse processing in MT: problems in pronominal translation”. *Proceedings of the 13th International Conference on Computational Linguistics (COLING'90)*, Helsinki, Finland.

Wakao, Takahiro. 1994. "Reference resolution using semantic patterns in Japanese newspaper articles". *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94)*, 1133-1137. Kyoto, Japan.

Walker, Marylin. 1989. “Evaluating discourse processing algorithms”. *Proceedings of the 27th Annual Meeting of the ACL (ACL'97*, 251-261. Vancouver, Canada.

Walker, Marylin, M. Iida & S. Cote . *Japanese discourse and the process of centering*. IRCS Report No. 92-14, The Institute for research in Cognitive Science, University of Pennsylvania.

Walker, Marylin, M. Iida & S. Cote. 1994. "Japanese discourse and the process of centering". *Computational Linguistics*, 20 (2).

Webber, Bonnie L. 1979. *A formal approach to discourse anaphora*. London: Garland Publishing.

Wilks, Yorick. 1973*. Preference seman*tics. Stanford AI Laboratory memo AIM-206. Stanford University.

Wilks, Yorick A. 1975. "Preference semantics". *The formal semantics of natural language* ed. by E. Keenan. Cambridge University Press.

Wilks, Yorick & Mark Stevenson. 1998. The Grammar of Sense: Using part-of-speech tags as a first step in semantic disambiguation. *Journal of Natural Language Engineering* 4(3) (forthcoming)

Williams, Sandra, Mark Harvey & Keith Preston. 1996. "Rule-based reference resolution for unrestricted text using part-of-speech tagging and noun phrase parsing". *Proceedings of the International Colloquium on Discourse Anaphora and Anaphora Resolution (DAARC)*, 441-456. Lancaster, UK.

Winograd, Terry. 1972.*Understanding natural language*. New York: Academic Press.

Weisweber, Wilhelm. “The experimental MT system of the project KIT-FAST - complementary research for Eurotra-D”. *Proceedings of the International Conference "Machine translation, 10 years on"*, 12.1-12-19. Cranfield, UK.