

ANAPHORA RESOLUTION AS LEXICAL COHESION IDENTIFICATION

Samuel W.K. Chan and Benjamin K. T'sou

Language Information Sciences Research Centre
City University of Hong Kong, Hong Kong

ABSTRACT

Anaphora, an important indicator in lexical cohesion, is a discourse level linguistic phenomenon. Most theoretical linguistic approaches to the interpretation of anaphoric expressions propose a treatment on the basis of purely syntactic information. In this article, what we proposed is to cast anaphora resolution as a semantic inference process in which combination of multiple strategies, each exploiting a different linguistic knowledge, is employed to resolve anaphora into a coherent one. We also exhibit how to embed an anaphora resolution into a framework which captures all the salient parameters as well as to remedy, to a certain extent, the inadequacies found in any monolithic resolution systems. The effectiveness of the anaphora resolution considered in this work is exemplified through a set of simulations.

1. INTRODUCTION

It is realized that pieces of information in a text are highly interconnected and links between them are mainly local, i.e., linguistic items connect adjacent or near adjacent clues. One of the principal devices indicating such interconnections is anaphora. Anaphora is the linguistic phenomenon of abbreviated subsequent reference. It is also a device for referring back to an entity which has been introduced with more fully descriptive phrasing earlier in the text. The entity may be an object, a concept, an individual, or state of being. In the linguistic studies of anaphora, although some comprehensive studies have already presented and a coherent classification of anaphora has been laid, much of them is still in a theoretical basis and no much computational algorithms are released [1,13,15]. Among a few, systematic methods have been proposed for resolving anaphora, however, as in work on other syntactic processing, most anaphora resolution algorithms are conceptualized entirely within symbolic frameworks [3,10,11,9]. Syntactic information plays a central role in establishing appropriate referents, nevertheless, it has long been recognized the traditional symbolic syntactic approach which tries to capture the meaning from antecedent linguistic items in text cannot be translated into a processing theory [2]. Anaphoric relations could be hidden in the context. Interpretation of anaphora is crucial in making use of syntactic and semantic preferences of each linguistic item in sentences, moreover, it is a great necessity to incorporate previously analyzed sentences which is regarded as an important knowledge source in modelling prior specific context effect.

Psycholinguistic accounts of discourse comprehension often assume anaphora resolution proceeds by *search-and-match*. Interpretation of an anaphora is not resolved by decoding, but inferred, by a non-demonstrative inference process – a process of hypothesis formation and evaluation – in which linguistic constituents and contextual assumptions determine the class of possible hypotheses. These hypotheses are then evaluated in the light of certain general principles which authors and readers are expected to obey. The semantic inferences during anaphora resolution are those based on easily available information and those required making statements in a text locally coherent. To acquire a correct interpretation of a sentence, the reader must first identify the general interpretation of the input sentence, and combine these with the contextual assumptions generated in the preceding sentences to obtain the contextual effects. These contextual effects will govern the possible contextual implications

for the anaphora. In other words, a necessary step towards identifying the appropriate referent would be to search for linguistic items that could serve as antecedents. By comparing the coreferential devices and the linguistic items as to a match of various factors, including syntactic, semantic and case-role constraints, the set of potential referents might be considerably narrowed down.

In this article, we address the issue by treating semantic ambiguities and anaphora resolution in a same conceptual framework. The principle idea is that on encountering an anaphoric device, some sort of backward search is initialized in the framework which integrates knowledge sources from various paradigms. A collective knowledge source is used to determine the antecedent from a set of antecedent candidates. Disparate linguistic sources with weight assignments are used to encode the different constraints that are relevant for the resolution. The remainder of this paper is organized as follows. Section 2 explores in depth the rationale behind our framework. Anaphora resolution as well as semantic disambiguation are discussed in Section 3. The modelling of context effects in anaphora resolution is also exhibited. The system is implemented and simulations have been done as shown in Section 4, followed by a conclusion.

2. OVERVIEW OF SYSTEM ARCHITECTURE

In most accounts of what a reader does during comprehending a discourse, it is commonly accepted that readers potentially generate a rich variety of inferences while they construct a situation model of what the text is about. It is undisputed that both linguistic knowledge and discourse context facilitate the final interpretation of a syntactically or semantic ambiguous sentence. Knowledge captures the dependencies between concepts and provides coherence to the representation of text. A resolution of linguistic information across sentences, achieved primarily by relevant linkage of readers' background knowledge, has to be made in order to achieve a sentence interpretation. This seems, however, to be significantly difficult to achieve under the constraints imposed by a single knowledge source. In the following, we will explain how our framework can make use of different, partial, and orthogonal sources of linguistic information in facilitating semantic inferences, particularly in anaphora resolution.

2.1 Outline of the Framework

The framework is a hybrid model which combines a syntactic representation of the input sentence, a set of lexical subsymbols and a connectionist constraint-satisfaction process. The framework consists of several major constituents: A proposition parser, a lexicon, a syntactic network supported by a set of lexical subsymbols and an anaphora tagger. In this article, our concentration is focused on the syntactic network as well as the anaphora tagger.

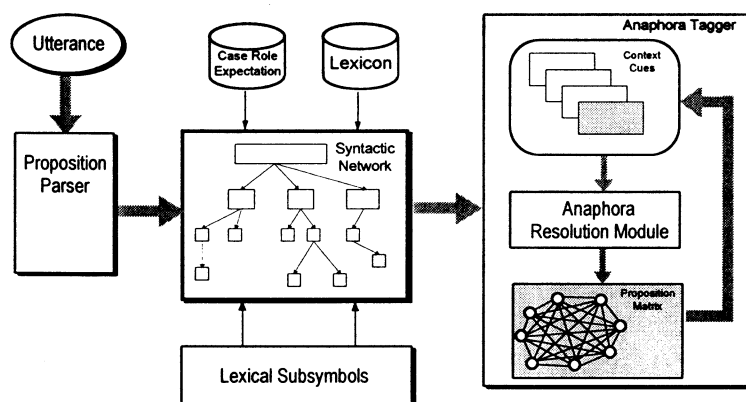


Figure 1: The System Framework

As shown in Figure 1, the input sentence is first parsed by a proposition parser which turns the sentence into a set of subpropositions. The resultant subpropositions are connected to each other as a

syntactic network which is modified from the Bayesian model [12]. The network which is based on the use of probabilistic causal relationships among the subpropositions is to model the uncertainties involved in syntactic inferences. The whole syntactic network is supported by a set of lexical subsymbols accommodating all the semantic feature dimensions of the relevant concepts. The subsymbols, encoding the knowledge associated with other concepts, are generated in a backpropagation based neural network architecture [4]. Every concept in the syntactic network is linked with its subsymbols which roughly correspond to reasoning at a subconceptual level. Anaphora is resolved in the anaphora resolution module of the tagger. Based on the context cues developed from prior sentences and the linguistic knowledge generated from the current one, the tagger sets up a series of nodes in the working memory of the anaphora resolution module. Each of these nodes represents one of the linguistic items retrieved from the syntactic network, lexical subsymbols and the relevant context cues. Each node has its own activation while connections among nodes have strength values. A semantic inference takes place in the working memory inside the module in order to resolve the anaphora. In other words, the computations involved consist of two major components: A list of recalled linguistic items and a knowledge matrix which represents the interconnections among them. The knowledge matrix is established after extracting the linguistic knowledge from the syntactic network and context cues from the prior sentences. A constraint-satisfaction process involves the spreading of activations through the knowledge matrix in anaphora resolution. As long as a linguistic item's activation level is greater than some minimum threshold value, then the item is considered as part of working memory and is maintained in the knowledge matrix. However, if the activation of an item is less than the amount required performing an inference task, then the item will be deallocated, producing a kind of forgetting by displacement. The spreading is continued until an asymptotic activation vector is obtained. When the anaphora resolution is complete, the working memory in the tagger consists not only of the strength values of its linguistic items, but also of their patterns of interrelations which reflect the likelihood of what the anaphora is. Further details of each basic component are discussed in the following sections.

2.2 Syntactic Network

Reasoning about knowledge plays a central role in linguistic inferences. It is obvious that many concepts in natural language are vague rather than precise. Our approach is to employ a Bayesian network to model the uncertainties involved in semantic ambiguity, particularly in anaphora resolution. A Bayesian network is a probabilistic model in which an acyclic directed graph is used to represent probabilistic causal relationships among random variables. It consists of nodes and links. Each node in the network represents a probabilistic variable as one of n random variables in the set $X = \{X_1, X_2, \dots, X_n\}$. Each link is a directed edge from X_i and X_j and indicates the existence of a direct causal influence between X_i and X_j . In particular, a node Y is said to be the parent node of X if there is a link from Y to X . Bayesian networks are quantified by both specifying the distribution of each variable in the network as well as the conditional probability matrices associated with the links. One use of Bayesian networks has been in the field of expert systems [6]. These systems often represent knowledge as *if-then* rules: *if A then B with certainty C*. In probabilistic terms, this would correspond to the conditional probability statement $P(B|A) = C$. Also, variables whose values have been observed, and are therefore known, are called evidence E , and can be regarded as constituting the input to the network and the degree of belief of other nodes can be calculated.

Figure 2 shows in more detail the nodes in the syntactic network in our application. Upon the initial analysis of the proposition parser, the formation of the syntactic network comprises both the identification of the corresponding subpropositions and the organization of these subpropositions into a hierarchy which is subject to various restrictions imposed by the grammar. For each input sentence, the parser defines a set of subpropositions and concept nodes, $X_i \in X$. From the standpoint of utilizing Bayesian networks as a modelling tool in anaphora resolution, what needs to be done is to specify the parent nodes $P(X_i)$, for each X_i in X and the conditional probability matrices, usually with non-negative entities, associated with the links. To be more precise, from the proposition parser, all the relevant subpropositions X_i are aggregated and an ordering is set up for all $X_i \in X$. Inspection is done to make sure whether there are any subpropositions left in X . In the affirmative case, a node X_i is

picked up and to be added as a node to the syntactic network. At the same time, the parent node $P(X_i)$ is being set up to some minimal set of nodes already in the net. The conditional probability matrix for X_i which reflects the knowledge from the lexicon is then defined.

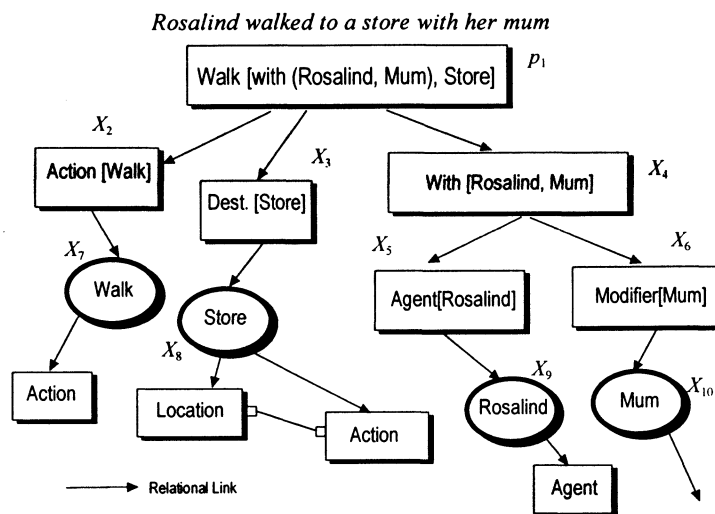


Figure 2: A fragment of syntactic network in representing a sentence, *Rosalind walked to a store with her mum.*

In addition, the semantic deep cases based on case grammars [8] are also represented in our syntactic network. There are restrictions on the semantic cases which each linguistic items may adopt and that are embedded within the linguistic items. These restrictions, in the form of probability matrices, are based on the occurrence of each concept in the subpropositions. For instance, the case of *store*, in Figure 2, can either be a *location* or an *action*, but not an *agent*. The possible case role nodes are allocated as the child nodes $C(X_i)$. Zero conditional probabilities are allowed to indicate the inhibitions between the nodes and they are denoted by the links with open square at both ends. For example, the activations of the case *location* will inhibit the case *action*, in order to ensure the non-coexistence of both concepts in the final interpretation. This whole process need not be comprehensive but just powerful enough to ensure some of the plausible nodes are activated. Finally, all the links are created on an empirical basis, following one's intuition about the domain being modelled.

Many problems in applied natural language processing hinge on relating lexical units to others that are similar in meaning. Our lexical subsymbols provide a similarity measure between every lexical units in order to capture lexical preference which is the key to resolve ambiguity in natural language understanding [16]. Taraban and McClelland also show that the structural models of language analysis are not in fact good predictors of human behaviour in resolving ambiguity [14]. Obviously, category formulation of lexical items is a fundamental problem in natural language understanding and it is also the basis of the formation of our lexical subsymbols. Basically, in our design, each lexical item has different senses (WS_i) and there are semantic relations between the senses (WS_i) and the concepts (C_i) which are used to define them. The sense WS_i has been broken up into components. This *componential* approach is the basis of our formation of lexical subsymbols which are used to capture the semantic feature dimensions of the relevant concepts. On the other hand, word sense definitions are marked with a subject field code (SF_i) which tell us which subject area the word-sense pertains to. For example, the *financial-sector-related* senses of *bank* are marked EC (*Economics*) as its subject field code with *money*, *deposit* and *repayment* as its subordinate concepts. The formation of the lexical subsymbols is on the basis of considering the subject fields in which sense definitions are commonly shared.

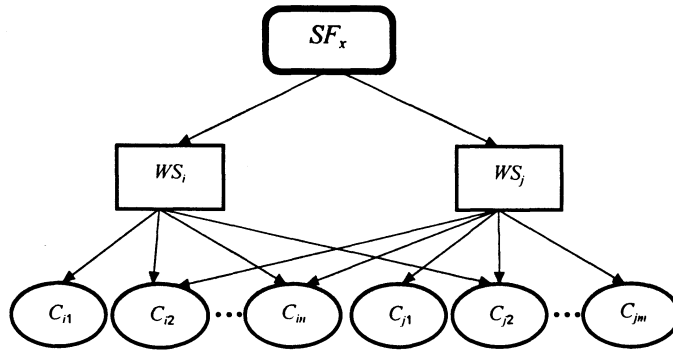


Figure 3: The relations between subject field code (SF), different word senses (WS) and their related concepts (C)

The subsymbols are formed through a backpropagation neural network architecture, however, the detail discussion of the neural architecture is out of the scope of this paper [7,4]. Figure 4 shows some of the lexical subsymbols. Each of the lexical subsymbols is a 16-dimension data vector, which has its distributed representation associated with its semantic meaning. They form granules that are the fuzzy sets of points having the form of a clump of elements drawn together by similarity. The lexical subsymbols capture the corresponding word senses from their relevant subordinate concepts and the related subject fields.

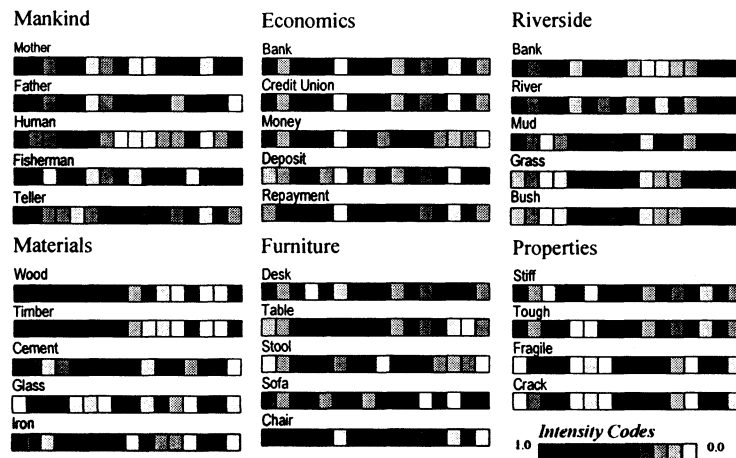


Figure 4: Sample lexical subsymbols

On the other hand, these lexical subsymbols allow a greater tolerance of errors in activation values. This is certainly not the case in conventional symbolic approaches to the representation of meanings. The semantic similarity measure between the lexical subsymbols is defined by

$$SS(x_i, x_j) := \begin{cases} \langle x_i, x_j \rangle & \text{if } \langle x_i, x_j \rangle \geq d_{max} \\ 0 & \text{otherwise} \end{cases} \quad (\text{Eqn. 1})$$

where $\langle x_i, x_j \rangle$ is the dot product of the lexical subsymbols x_i, x_j and d_{max} is proportional to the number of lexical subsymbols in the system. It is most appropriate to regard the lexical subsymbols as sets of cues that constrain the meanings of the related lexical items. It is conceivable that lexical items with similar meanings are represented using profiles which are similar but not identical, such as the lexical subsymbols of *Wood* and *Timber*, *Bank* and *Credit Union*, but not the *Banks* in different subject codes. As a result, a whole spectrum of lexical similarity and cohesion can be captured. In short, syntactic reasoning is hard to implement computationally because it leads to a blowout of computational complexity and also requires knowledge of many probabilities that in fact one doesn't know. Our lexical subsymbols provide the complement nicely, and have proved sophisticated in many areas where depth of reasoning is not required, in particular in anaphora resolution. These lexical

subsymbols serve as a means to reflect the lexical cohesion which is the result of a chain of related words that contribute to the continuity of lexical meaning. The lexical cohesion, as shown in the simulations in Section 4, provides an easy-to-determine context to aid in the resolution of anaphora and in narrowing down the search space to a specific meaning.

3. ANAPHORA RESOLUTION

It has been pointed out in the last section that the syntactic network supported with lexical subsymbols gives a rough sketch of the input. The process needs not be comprehensive but just powerful enough to ensure that some of the plausible propositions or lexical items are activated. In this section, we report on the application of an anaphora resolution to allow all senses to be determined at once in a computationally effective way. The resolution need to be involved two main ingredients. First, we activate the relating linguistic items from the syntactic network for each linguistic input and select a small number of its most closely associated lexical subsymbols. It involves calculating the inter-relationships between the extracted subpropositions, the possible case-roles, and the relevant lexical subsymbols. They illustrate both the syntactic and semantic cues of the current sentence. At the same time, context cues from prior sentences are also incorporated in order to make the sentences as a coherent whole. Details of the anaphora resolution are further described in the following.

3.1 Reasoning in Anaphora Resolution

The first step involved in the anaphora resolution is to construct an associative network in order to represent the interrelations among the extracted linguistic items. Let K be a knowledge matrix and $K(i,j)$ denotes the interrelationship between the linguistic items i and j in the associative network. The formation of knowledge matrix K is achieved first by propagating the initial evidence vector, which is equal to one, from the input proposition α of the syntactic network down to its lexical nodes. For each extracted linguistic item X_i of the input proposition α , a corresponding initial evidence vector $(e_1, e_2, \dots, e_n) \in E$ is then initialized, i.e., if $X_i \in \alpha$, e_i is then assigned to be unity. On the other hand, if X_i is only an element in the syntactic network, e_i is assigned to the degree of belief while all others are set to be zero. Each evidence vector serves as independent evidence. At the same time, for each extracted linguistic concept X_i , activate some of the highly associative lexical subsymbols. For each evidence vector, propagate the evidence through the syntactic network. If the maximal implication chains have been traversed, the resultant vector is assigned as the corresponding column of the knowledge matrix K . The steps are repeated until all columns of the knowledge matrix are completed. The knowledge matrix K of an input sentence represents all the relevant disparate sources of sentential information, which obviously includes syntactic restrictions, case-role expectations, lexical associations and some semantic rules. These inter-relationships between the linguistic items, represented in the knowledge matrix K , need not be very sophisticated. They are to be just powerful enough so that the right linguistic items are likely to be among those generated, even though others will also be generated that are irrelevant or even outright inappropriate.

In fact, what the knowledge matrix represents is the preferences which candidate referents are most preferred. Instead of considering a single candidate referent for an anaphora and ruling out any possible candidates based on syntactic structures as in other systems, our framework elicits all the possible candidates with an individual weight assignment. Case roles and lexical subsymbols may also cast votes for those possible candidates which are most closely matched to the anaphora. In effect, all the knowledge sources indicate a possible preference, or a set of preferences, of the candidate referent in the absence of any hard constraints. Certainly, it is not difficult to understand the mechanism described above uses weak, but robust construction strategies, even though some of them may be self-contradicted. The facts drawn from above are classified as a weak production system and the whole process seems lack of guidance and intelligence. It simply produces potential inferences, in the hope that some of them might turn out to be true. The knowledge matrix is an initial, enriched, but incoherent and possibly contradictory representation which is then subjected to a constraint satisfaction process in order to integrate the meanings into a coherent whole. The

mechanism of the resolution process is to propagate a state vector $U = (u_1, u_2, \dots, u_n)$, where u_i is the activation level for the element i initially to the knowledge matrix K such that $U(t+1) = T[U(t)K]$ where $U(t)$ is the activation vector at some discrete time t , and T is the vector normalising operation. The process stops at iteration m if $|U(m) - U(m-1)| < t$, where t , the tolerance, is another preset threshold which is used to control the number of inhibited elements in the process. Mathematically, the constraint satisfaction process is a vector-matrix multiplication. It can be regarded as a Markov process with modification of the transition probability by repeated vector multiplication which leads to equilibration [5]. The final activation vector $U(m)$ shows how strongly related items have strengthened each other, while unrelated or contradictory items have low activation value. The process strengthens the contextually appropriate linguistic items and inhibits unrelated and inappropriate ones, so that smart and complex deductions can be achieved. In other words, the resolution is used to reduce the dimensionality of the extracted linguistic items and exclude unwanted from the knowledge network. As a result, the most preferred candidate referent is unified under the restriction of all the possible knowledge sources.

3.2 Context Cues in Anaphora Resolution

Throughout the literature of semantic understanding runs a common theme: Language is understood in context. Linguists appeal to immediate linguistic context to give an account of interpretation of ambiguous expressions. The only inferences coded automatically during interpretation are those based on easily available information and those required making sentences in a text locally coherent. To acquire a correct interpretation of an anaphora, the reader must first identify the general interpretation of the input sentence, and combine these with the contextual assumptions generated in the preceding sentences to obtain the context effects. These context effects will govern the possible contextual implications for the anaphora. As can be seen in the forgoing sections, for each input sentence after the resolution process, our framework generates a vector representation that exhibits all the remaining linguistic items. They are contextually relevant and all conflicts, as well as irrelevancies, have already been eliminated. To accomplish our objective of modelling context effects in anaphora resolution, we have adopted a *proposition matrix* Γ in which linguistic items are encoded and stored as patterns of interconnections between the items. More specifically, the linguistic items form a single, common proposition matrix which stores information about individual linguistic items, as well as associative information about pairs of items. Rows and columns of the proposition matrix correspond to the distilled linguistic elements and the connections between them are represented by the nonzero entries. The matrix also specifies how strongly each distilled linguistic item in the sentence is related to every other.

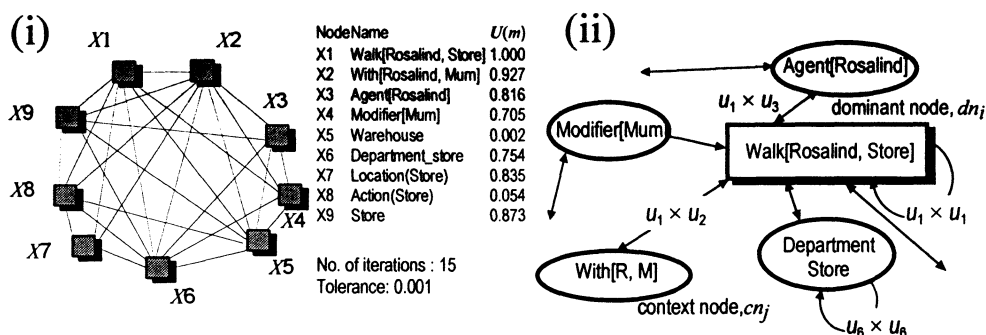


Figure 5: (i) After reading the first sentence, *Rosalind walked to a store with her mum. She brought a dress*, a fully knowledge matrix K with rough, or even outright contradictory nodes is formed and subjected to the resolution process. The top right column shows the asymptotic activation after the process. It is apparent that the incorrect meaning of *Store*, *Warehouse*, is deactivated. (ii) After the resolution, a proposition matrix Γ having the distilled linguistic elements is formed which may be classified into dominant nodes or context nodes. This matrix captures the context effect from the first sentence and is again an important knowledge source to resolve the pronoun *She* in the second sentence.

As shown in Figure 5, the proposition matrix Γ is constructed from the asymptotic activation of each element. It represents the interrelations between elements remaining after the anaphora resolution process. The strength of connection between linguistic items is defined as $\Gamma(r, s) = u_r \times u_s$, where u_r, u_s are the final asymptotic activation values of the r th and s th linguistic items respectively. A given proposition matrix is a complex collection of information that contains the *distilled* linguistic items, such as lexical, syntactic and semantic, and associative information. The diagonal values of Γ represent the strength of linguistic elements in the matrix and the off-diagonal elements represent the strength of the relations between any two linguistic items. The item with largest strength is said to be the *dominant node* of the sentence while the others are called the *context nodes*. Figure 5(ii) shows the corresponding dominant node and context nodes. The proposition, *Walk[Rosalind, Store]*, is the dominant linguistic node with the largest strength, $u_r \times u_s$, for the sentence. Figure 5(ii) also illustrates the connection strengths between the dominant node and its context nodes and demonstrates the computation of the connection strengths in the sentence. For each input sentence, the proposition matrix Γ summarizes the interrelations among the distilled elements constructed in the sentence. Obviously, this matrix captures a *slice* of context effect from the current sentence and is again an important knowledge source to resolve any anaphoric expression in any subsequent sentences in the paragraph. Based on the proposition matrices which reflect the interrelationships of the linguistic items from all prior sentences, the context effect on sentence j can be modelled mathematically in a context matrix M_j by the sum of k proposition matrices, such as

$$M_j = \alpha_1 \Gamma_{j-1} + \alpha_2 \Gamma_{j-2} + \dots + \alpha_k \Gamma_{j-k} \quad (\text{Eqn. 2})$$

In other words, the context effect is defined in a context matrix M_j which is constructed by adding the corresponding element from a set of k most adjacent proposition matrices. The feature gives an analytical simplicity in incorporating context cues into our anaphora resolution since the context matrix M_j and the knowledge matrix K are in the same morphological structure and associative recall can be represented as the element-by-element matching. It also indicates that as k increases, there is “relatively few” nonzero elements in the context matrix M_j . We formulate our discussion in applying the matrix in modelling the context effect as follows.

Definition 1. A subset of a context matrix M_j is called a *context cue* Q if it satisfies the following three conditions:

- (i) The subset contains at least three linguistic items.
- (ii) For each pair of linguistic items P and P , in the subset, both $P, \rightarrow P$, and $P, \rightarrow P$, are true.
- (iii) The subset is as large as possible; that is, it is not possible to add another linguistic items to the subset and still satisfy condition (ii).

The definition suggests that context cue Q is a full maximal subset which is self cohesive within the context matrix, even though M_j is sparse. This cohesive cue hints a group of inter-related semantic linguistic items in the forgoing sentences. It also provides a set of associations in the context which the reader can utilize to make the text hold together. The relevant context cue, as a set of traces, is incorporated in semantic interpretation by superimposing Q into the current knowledge matrix K_j to form a new knowledge matrix K_j^* as follows.

$$K_j^* = K_j \oplus Q$$

where $Z = X \oplus Y$ iff $z_{ij} = \begin{cases} x_{ij} + y_{ij} & \text{if } i, j \text{ are the common concepts in } X \text{ \& } Y \\ x_{ij} & \text{or otherwise} \end{cases}$

One of the most important implications in the application of context cues in the knowledge matrix is that all the previously analyzed sentences behave like a group of experts in the interpretation of the ambiguities in the current sentence. Each of the proposition matrices inherits the factual and circumstantial facts from the corresponding sentence. They encode some of the context knowledge which can be superimposed into the current knowledge matrix and carried over into the analysis as described. Obviously, the resultant knowledge matrix is more robust than an individual knowledge matrix because the linguistic cue is derived from a multiplicity of sources from the analyzed

sentences, making the links in the current knowledge matrix less prone to error. In general, context effects may be drawn from a wide variety of information sources that may include the preceding text, such as the cultural or scientific knowledge, item of shared or idiosyncratic information that the hearer has access to at the time. It is worthwhile to mention that, in this framework, we limit our scope of context effect by simply considering the preceding linguistic text, or the set of assumptions in which the utterance takes place.

4. SIMULATION EXPERIMENTS

This section describes a step-by-step operation of the above mentioned mechanism, concentrating on the effect of various constraints in accounting anaphora resolution, to supplement, verify and strengthen our theoretical considerations above. While a number of diverse theories have been proposed for certain classes of anaphors, such as centering, reference resolution remains a difficult problem and continues to be a major hurdle on the critical path to language understanding. As mentioned before, there is a strong tendency in syntactic approach for determining the referent of a pronoun. However, structure does not completely determine the issue. Examples that cannot be resolved on syntactic basis and require general knowledge reasoning are commonly found. Let us consider the following two sentences,

(S-1a) *Susan dropped the glass on the table. It broke.*

(S-1b) *Susan dropped the table on the glass. It broke.*

The pronoun *it* appears to refer to the *glass* in both cases. Determine the correct referent in both of these sentences can only be a result of general knowledge about *table* and *glass*. In our framework, the full syntactic decomposition of the sentence (S-1a) produced by the proposition parser is:

```
% Prop_Parse(S-1a)
Drop[Susan, Glass, Table], Action[Drop], Agent[Susan],
Object[Glass], Break[Susan], Break[Glass], Break[Table], Action[Break],
Subject[Susan], Subject[Table], Subject[Glass],
Susan, Glass, Table, Break, Drop
```

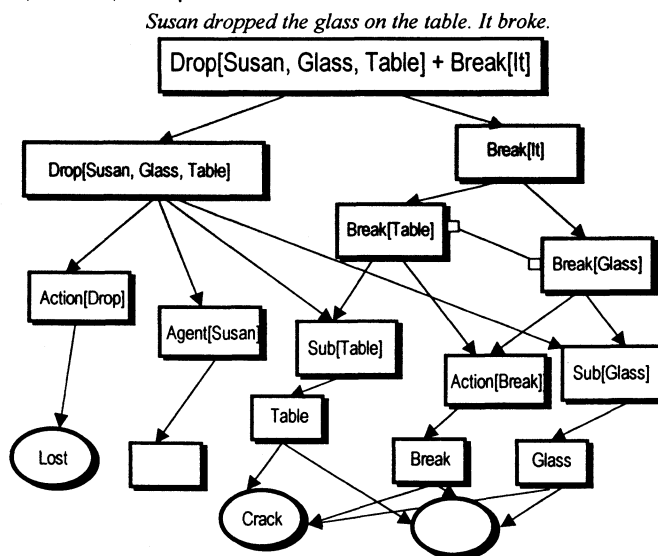


Figure 6: A fragment of syntactic network for the anaphora resolution in (S-1a), *Susan dropped the glass on the table. It broke.*

Obviously, anaphors do not introduce new entities into the sentence since they refer to prior entities. In our approach, each unique entity introduced into the text preceding an anaphora is assigned to be a potential candidate for resolving that anaphora, such as the three alternatives *Break[Susan]*, *Break[Glass]*, *Break[Table]*, for *it* in (S-1a). Resolving an anaphoric reference, in our framework,

requires choosing the best candidate among them. An initial knowledge matrix is first constructed by taking the current sentence with the anaphora and the two most recent possible candidates with their relevant context cues. The knowledge matrix is then subjected to the anaphora resolution which indicates how cohesive will the candidates blend with the current anaphora by inspecting its activation change. The “losing” candidate is discarded, and another knowledge matrix is constructed by pairing the “winner” with the next candidate. This process continues until every candidate has been examined. The last winner is then chosen as the anaphor’s referent. For simplicity, in the anaphora resolution of (S-1a), the only competition between Table and Glass is considered.

Figure 6 shows the syntactic network of the input sentences. Both of the possible candidate for the pronoun *it* are activated in parallel, and concept nodes are added. To simplify the simulation, all the elaborations are ignored and only the most essential nodes are displayed. The two potential candidates are connected by inhibitory links, as represented in open squares at both ends, to ensure their non-coexistence in the final interpretation. As shown in Figure 4, buoyed by the closed associations and high similarity measures between *Fragile*, *Glass* and *Crack* in terms of lexical subsymbols, this leads to the dominant role of the linguistic item *Sub[Glass]* and its ancestor, *Break[Glass]*. At the same time, the noncoexistence link between the *Break[Glass]* and *Break[Table]* deactivates the subtree of *Break[Table]*. Figure 7 shows the corresponding knowledge matrix so formed. The initial activation of each node is shown as $U(0)$ in Figure 7.

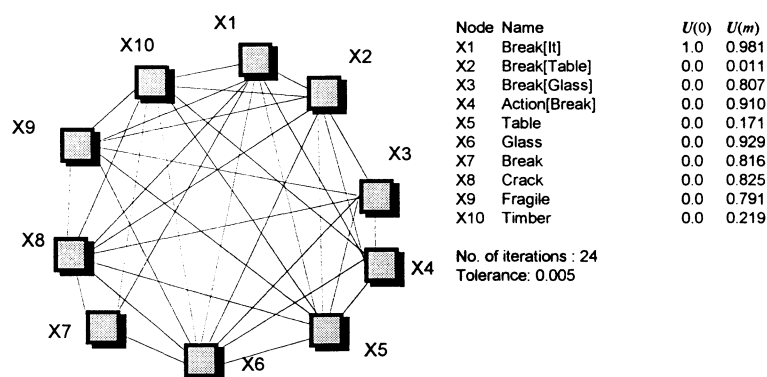


Figure 7: The resulting knowledge matrix in the anaphora resolution *it* in the sentence (S-1a).

After twenty-four iterations, by reading the final activation of each linguistic item in the knowledge matrix as shown in the rightmost column of Figure 7, *Break[it]* is unified with *Glass* in the competition. The anaphora resolution process deactivates the inappropriate linguistic items. The three nodes representing the incorrect candidate pronoun *it*, *Break[Table]*, *Table* and *Timber* are deactivated. Our system resolves the anaphora by determining which are more likely to break in certain situations through the world knowledge encoded in the lexical subsymbols. The top rightmost column in Figure 7 indicates the asymptotic activation vector in the anaphora resolution of the sentence (S-1a). The pronoun *it* in sentence (S-1b) will also be resolved in a similar manner. This demonstrates that our framework has some capabilities in anaphora resolution, even though there is no much syntactic cues, by taking into account both world and linguistic knowledge and without demanding either kind of knowledge be dominant. In another example, let us consider the pronoun resolution in following text.

[Judy]₁ is going to have [a birthday party]₂. [She]₃ wants a [hammer]₄ for [present]₅.
Then [she]₆ can fix her [coat rack]₇. [She]₈ asks for [her mother]₉ to get [it]₁₀ for
[her]₁₁. [Her mum]₁₂ thinks [girls]₁₃ should not play with [a hammer]₁₄. [She]₁₅ buys
[her]₁₆ [a dress]₁₇.

This fragment contains a total of seventeen simple discourse entities. Entities 3, 6, 8, 10, 11, 15 and 16 are pronouns. In the upcoming demonstration, we try to resolve the problem of pronoun referents, the entity 10 as being *a hammer*. As discussed above, our anaphora resolution is actually composed of pairs constructed by grouping the current analyzed sentence with the other potential candidates. Resolving an anaphoric reference requires choosing the last surviving candidate in the resolution. It is

patently infeasible for all previously introduced candidates are examined for each encountered anaphora. In particular, since the most correct candidate is likely to be fairly close to the anaphora, the number of comparison is restricted to an arbitrary number in order to reduce the computational complexity. Furthermore, candidates are presented to the resolution in a linear basis, i.e., from the most closest to farthest. Figure 8 shows the knowledge matrix so formed in identifying the entity 10.

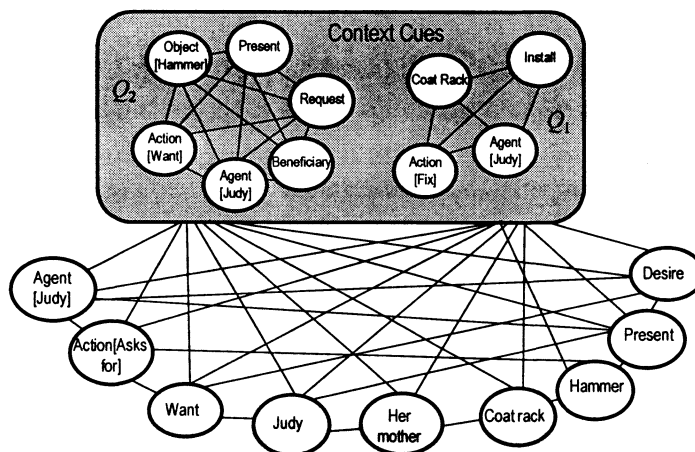


Figure 8: A knowledge matrix of the sentence. $[She]_8$ asks for $[her\ mother]_9$ to get $[it]_{10}$ for $[her]_{11}$ with the two possible candidates of pronoun *it*.

The most distinctive property of resolution is the involvement of buffers which capture context cues and are designed to carry each preceding analyzed knowledge sources over into the current processing cycle, in the hope that they will serve as common bridging elements between the sentences. In order to avoid unnecessary generation of alternatives as most syntactic systems tend to have, we limit our resolution process of entity 10 with the following possible candidates: $[hammer]_4$, $[present]_5$, and $[coat\ rack]_7$. As shown in Figure 8, initial hypotheses are constructed by taking the anaphora *it* and the three possible candidates. At the same time, the context cues Q_1 and Q_2 are linked with the corresponding concepts in the sentence in the buffer. The degree of lexical cohesion and repetition between the current sentence with the context cues allow the system to identify Q_2 that is relevant to the current sentence. Due to the association of the linguistic items *Action[Want]*, *Action[Asks for]*, and the co-occurrence of *Agent[Judy]* in the sentence and Q_2 , the system resolves *it* as *hammer* by virtue of the fact that *hammer* is the object of the *Action[Want]* in the context cue.

The simulations presented here establish a framework of anaphora resolution as a candidate for a viable theory of sentence processing. The weakly interactive framework is used to model various types of semantic inferences. Context effects in the anaphora resolution have also been explained. This model has exhibited its relative capabilities in language interpretation.

5. CONCLUSIONS

According to modern trends in linguistics, text rather than sentences are the units of communication. There is a lot of ongoing research in finding textual regularities on the basis of empirical investigations. Among many factors that furnish discourse coherence, lexical cohesion has been identified to be one of the most important features that characterizes the structure of a text, distinguishing it from a non-text. Anaphora, in conjunction with other lexical repetition, contributes the textual coherence by creating cohesive ties within the text. In this paper, we have demonstrated and simulated a framework for anaphora resolution. Treating anaphoric referencing as a resolution of collective evidences, the framework makes use of both syntactic surface structures and semantic associations in their lexical subsymbols. A mechanism which underlies human apparent abilities to process discourses in both integrative and incremental way is also demonstrated.

ACKNOWLEDGMENTS

We would like to thank the two anonymous reviewers for their helpful comments. This research was supported in part by the Strategic Research Grant of City University of Hong Kong under grant # 7000868.

REFERENCES

- [1] Ariel, M. (1994). Interpreting anaphoric expressions: A cognitive versus a pragmatic approach. *Journal of Linguistics*, 30, 3-42.
- [2] Brown, G. & Yule, G. (1983). *Discourse Analysis*. Cambridge University Press.
- [3] Carter, D.M. (1987). *Interpretation Anaphora in Natural Language Texts*. Ellis Horwood.
- [4] Chan, S.W.K. (1995). Inferences in natural language understanding. *Proceedings of the Fourth IEEE International Conference on Fuzzy Systems*, vol.2, 935-940. Yokohama.
- [5] Chan, S.W.K., & Franklin, J. (1998). Symbolic connectionism in natural language disambiguation. *IEEE Transactions on Neural Networks: Special Issue on Neural Networks and Hybrid Intelligent Models: Foundations, Theory, and Applications*, 9(5), 739-755.
- [6] Cooper, G.F. (1989). Current research directions in the development of expert systems based on belief networks. *Applied Stochastic Models and Data Analysis*, 5, 39-52.
- [7] Dyer, M.G., Flowers, M., & Wang, Y.J.A. (1992). Distributed symbol discovery through symbol recirculation: Toward natural language processing in a distributed connectionist network. In Reilly, R.G. & Sharkey, N.E. (Eds.), *Connectionist Approaches to Natural Language Processing*, 21-48, Lawrence Erlbaum Associates.
- [8] Fillmore, C.F. (1968). The case for case. In E. Bach & R.T. Harms (Eds.), *Universals in Linguistic Theory*, New York, 1-90.
- [9] Lappin, S., & Leass, H.J. (1994). An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20, 4, 535-561.
- [10] Mitkov, R. (1994). An integrating model for anaphora resolution. *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94)*, Japan.
- [11] Nasukawa, T. (1994). Robust method of pronoun resolution using full-text information. *Proceedings of the 15th International Conference on Computational Linguistics (COLING'94)*, Japan.
- [12] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
- [13] Sag, I.A., & Hankamer, J. (1984). Toward a theory of anaphora processing. *Linguistics and Philosophy*, 7, 325-345.
- [14] Taraban, R., & McClelland, J.L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectation. *Journal of Memory and Language*, 27, 597-632.
- [15] van Hoek, K. (1997). *Anaphora and Conceptual Structure*. University of Chicago Press.
- [16] Whittemore, G., Ferrara, K., & Brunner, H. (1990). Empirical study of predictive powers of simple attachment schemes for post-modifier prepositional phrases. *Proceedings of 28th Annual Meeting of the Association for Computational Linguistics*, 23-30.