

# A Sequenced Model of Anaphora and Ellipsis Resolution

Shalom Lappin  
Department of Computer Science  
King's College, London  
lappin@dcs.kcl.ac.uk

June 27, 2003

## Abstract

I compare several types of knowledge-based and knowledge-poor approaches to anaphora and ellipsis resolution. The former are able to capture fine-grained distinctions that depend on lexical meaning and real world knowledge, but they are generally not robust. The latter show considerable promise for yielding wide coverage systems. However, they consistently miss a small but significant subset of cases that are not accessible to rough-grained techniques of interpretation. I propose a sequenced model which first applies the most computationally efficient and inexpensive methods to resolution and then progresses successively to more costly techniques to deal with cases not handled by previous modules. Confidence measures evaluate the judgements of each component in order to determine which instances of anaphora or ellipsis are to be passed on to the next, more fine-grained subsystem.<sup>1</sup>

## 1 Introduction

Anaphora and ellipsis resolution have been an important focus for work in natural language processing over the past twenty-five years. Providing adequate solutions to these tasks is necessary for the development of genuinely

---

<sup>1</sup>Earlier versions of this paper were presented at the 4th Discourse Anaphora and Anaphora Resolution Colloquium in Lisbon in September, 2002, the Linguistics Colloquium at the University of Toronto in November, 2002, and the Linguistics Colloquium at the University of Reading in January, 2003. I am grateful to the audiences of these forums for useful discussion of the ideas presented here. I would also like to thank Ruslan Mitkov and Andy Kehler for their encouragement and their helpful comments on this work.

robust systems for (among other applications) text interpretation, dialogue management, query answering, and machine translation. A wide variety of methods have been applied to the treatment of anaphora and ellipsis ranging from knowledge intensive and inference-based techniques to statistical modelling and machine learning. In this paper I will provide an overview of the main approaches and summarize their comparative strengths and limitations. My concern in this survey is not to offer a detailed account of the numerous computational treatments of anaphora and ellipsis that appear in the literature but to indicate the main advantages and shortcomings of the primary approaches that have been suggested.<sup>2</sup>

I will then sketch an integrated model which employs alternative techniques in a sequence of ascending computational cost and domain specificity. This model first invokes relatively inexpensive wide coverage procedures for selecting an antecedent for a pronoun or an elided element. It then moves through successively more expensive, fine grained measures to handle the cases not resolved by the preceding modules. It applies confidence measures to the decisions of each module to evaluate the reliability of its output. In this way it determines, for each module, which cases have been correctly resolved and which ones are passed on to the following component.

In Section 2 I look at knowledge-based and inference driven approaches to pronominal anaphora resolution. Section 3 considers various knowledge-poor methods for anaphora interpretation. Section 4 extends the comparison to VP ellipsis, and Section 5 takes up fragment interpretation in dialogue viewed as a type of ellipsis. Finally, in Section 6 I describe the proposed sequenced model. Section 7 states conclusions and indicates directions for future work.

## 2 Knowledge-Based and Inference-Driven Approaches to Anaphora

Knowledge-based approaches to anaphora resolution generally rely on rules of inference that encode semantic and real world information in order to identify the most likely antecedent candidate of a pronoun in discourse. An interesting example of this approach is Kehler's ((Kehler, 2000), (Kehler, 2002)) use of (Hobbs, Stickel, Appelt, & Martin, 1993)'s model of abductive

---

<sup>2</sup>See (Mitkov, 2002) for a recent study of anaphora resolution that includes a history of the problem within natural language processing. See (Mitkov, Boguraev, & Lappin, 2001) for examples of current work on anaphora resolution. (Huang, 2000) offers an extensive cross linguistic investigation of anaphora and examines alternative linguistic theories of this relation. See (Lappin, 1996) and (Lappin & Benmamoun, 1999) for theoretical and computational discussions of ellipsis resolution.

reasoning to establish coherence relations among the sentences of a text. In Kehler’s theory pronouns are assigned antecedents through the abductive inference chains required for text coherence. (Hobbs & Kehler, 1997) and (Kehler, 2002) also invoke abductive inference to interpret elided VP’s and resolve pronouns within VP ellipsis.

To illustrate this approach consider 1, to which (Kehler, 2000) assigns the representation 2

(1) John hid Bill’s keys. He was drunk.

(2) a.  $hide(e_1, John, Bill, ck) \wedge car\_keys(ck, Bill)$   
 b.  $drunk(e_2, he)$

He uses axioms like those in 3 to construct the backwards abductive inference chain in 4 from 2 to a conclusion in which *he* is resolved to *Bill* (4g).

(3) a.  $\forall e_i, e_j (cause(e_j, e_i) \Rightarrow explanation(e_i, e_j))$   
 b.  $\forall x, y, e_i (drunk(e_i, x) \Rightarrow \exists e_j, e_k (not\_want(e_j, y, e_k) \wedge drive(e_k, x), \wedge cause(e_i, e_j)))$

(4) a.  $explanation(e_1, e_2)$   
 b.  $cause(e_2, e_1)$   
 c.  $cause(e_2, e_3 \wedge cause(e_3, e_1))$   
 d.  $cause(e_2, e_4 \wedge cause(e_4, e_3))$   
 e.  $not\_want(e_3, John, e_5) \wedge have(e_5, Bill, ck)$   
 f.  $not\_want(e_4, John, e_6) \wedge drive(e_6, Bill)$   
 g.  $drunk(e_2, Bill)$

The main strength of knowledge-based systems is their capacity to capture fine-grained semantic and pragmatic distinctions not encoded in syntactic features or frequency of co-occurrence patterns. These distinctions are not accessible to knowledge-poor approaches. They are crucial to correctly resolving pronominal anaphora and VP ellipsis in a small but important set of cases that arise in text and dialogue.

The two main difficulties with these systems are that (i) they require a large data-base of axioms encoding real world knowledge, and (ii) they apply defeasible inference rules which produce combinatorial blow up very quickly. Assigning cost values to inference rules and invoking a cost driven preference system for applying these rules (as in (Hobbs et al., 1993)) may reduce the blow up to some extent, but the problem remains significant.

As a result knowledge-based models of anaphora resolution are generally not robust. Their rules are often domain-dependent and hard to formulate in a way that will support inference over more than a small number of cases. Moreover, the semantic/discourse representations to which the inference rules apply are not reliably generated for large texts.

### 3 Knowledge-Poor Approaches

Knowledge-poor systems of anaphora resolution rely on features of the input which can be identified without reference to deep semantic information or detailed real world knowledge. One version of this approach employs syntactic structure and grammatical roles to compute the relative salience of candidate antecedents. Another uses machine learning strategies to evaluate the probability of alternative pronoun-antecedent pairings by training on large corpora in which antecedent links are marked.

(Hobbs, 1978) suggests one of the first instances of a syntactic salience procedure for resolving pronouns. He formulates a tree search algorithm that uses syntactic configuration and sequential ordering to select NP antecedents of pronouns through left-right, breadth-first traversal of a tree. (Lappin & Leass, 1994) propose an algorithm which relies on weighted syntactic measures of salience and recency to rank a filtered set of NP candidates. This algorithm applies to full syntactic parses. (Kennedy & Boguraev, 1996), (Mitkov, 1998), and (Stuckardt, 2001) modify and extend this approach to yield results for partial syntactic representations rather than full and unambiguous parse structures. (Grosz, Joshi, & Weinstein, 1995) employ a grammatical role hierarchy and preference rules for managing informational state change to select the local NP center (focus) for each element of the sequence of sentences in a discourse.

A recent instance of the application of machine learning to anaphora is (Soon, Ng, & Lim, 2001). They describe a procedure for training a classifier on a corpus annotated with coreference chains, where the NP elements of these chains are assigned a set of features. The classifier goes through all pairs of referential NP's in a text to identify a subset of coreferential pairs.

The obvious advantage of knowledge-poor systems relative to knowledge-

based models is that the former are computationally inexpensive and potentially robust. However, these claims of resource efficiency and wide coverage must be qualified by recognition of the expense involved in generating accurate syntactic representations for systems that apply to full parses or detailed grammatical role information. Saliency-driven systems also require domain specific and, possibly, language specific values for syntactic saliency measures. In the case of machine learning techniques it is necessary to factor in the cost of annotating large corpora and training classification procedures.

An important weakness of these models is that they cannot handle a small but significant core of anaphora resolution cases in which saliency cannot be identified solely on the basis of syntactic and morphological properties, and relative recency. These features are also the basis for the candidate rankings that machine learning methods generate.

(Dagan, Justeson, Lappin, Leass, & Ribak, 1995) attempt to enrich a syntactic saliency system by modelling (a certain amount of) semantic and real world information cheaply. They combine the Lappin-Leass algorithm with a statistically trained lexical co-occurrence preference module. Elements of the candidate antecedent list are assigned both saliency and lexical preference scores. The latter are based on frequency counts for verb-NP and prep-NP pairs in a corpus, and the substitution of the candidate for the pronoun in the observed head-argument relation of the pronoun. When the difference between the saliency scores of the two highest ranked candidates is below a (experimentally) determined threshold and the lexical preference score of another candidate  $C_i$  exceeds that of the first by a (experimentally) specified ratio, then  $C_i$  is selected.

Consider the pronoun *it* in 5.

- (5) The utility (CDVU) shows you a LIST4250, LIST38PP, or LIST3820 file on your terminal for a format similar to that in which *it* will be printed.

The statistical preference module overrides the higher syntactic saliency ranking of *utility* to select *file* as the antecedent of *it*. This preference is due to the fact that *print file* has a significantly higher frequency count than *print utility*. The statistical module improved the performance of Lappin and Leass's (1994) syntactic saliency-based algorithm from 86.1 % to 88.6 % on a blind test of 360 pronoun cases in a set of sentences taken from a corpus of computer manuals.

However, there are cases which still resist resolution even under the the finer grain of lexical co-occurrence information that such a statistical preference module provides. The contrast between 6 (= 1) and 7 illustrates

the limits of syntactic salience enriched with a statistically trained lexical preference metric.

(6) John hid Bill's keys. He was drunk.

(7) John hid Bill's keys. He was playing a joke on him.

*John* receives the highest syntactic salience ranking in both 6 and 7. Lexical preference conditions do not select between *John* and *Bill* in these cases. Reliance on real world knowledge and inference are needed to identify *Bill* as the antecedent of *he* in 6, and *John* and *Bill* as the antecedents of *he* and *him*, respectively, in 7.

## 4 VP Ellipsis

(Asher, Hardt, & Busquets, 2001) specify a knowledge-based approach to the interpretation of VP ellipsis. They employ a general parallelism constraint based on (Asher, 1993)'s notion of *Maximal Common Theme* (MCT) to resolve ambiguity in VP ellipsis. They define a *Theme* for a Discourse Representation Structure (DRS)  $K$  as a DRS  $K'$  obtained from  $K$  by the application of 0 or more operations of a certain type on  $K$ . These operations are (i) deletion of a discourse marker, (ii) deletion of an atomic condition, and (iii) systematic renaming of a bound discourse marker. A *Common Theme* (CT)  $T$  of two DRS's  $J$  and  $K$  is a theme of  $J$  and a theme of  $K$  which is such that for any other theme  $T'$  of  $J$  and  $K$ ,  $T$  is a theme of  $T'$ . Asher and Hardt's maximalization constraint states that in resolving scope ambiguity within a VP ellipsis construction, the preferred DRS for the elided VP and its antecedent is the DRS that provides the MCT for the DRS's representing each clausal constituent. This constraint effectively constitutes a unification principle for the discourse representations of the sentences containing the elided and antecedent VP's.<sup>3</sup>

The MCT condition selects the parallel wide scope reading of the quantified NP *every student* relative to *a test* in 8.

(8) John gave every student a test, and Bill did too.

This is because the DRS's corresponding to this reading of each clause yield a CT that is a theme of the DRS's for the wide scope interpretation of *a test* relative to *every student*. The DRS's of the wide scope reading for *a*

---

<sup>3</sup>(Asher et al., 2001) also invoke this condition to resolve pronouns in ambiguous elided VP's.

*test* do not produce a theme for the DRS's of the wide scope reading of *every student*.

Several other instances of knowledge-based and inference-driven models of VP ellipsis interpretation are as follows. (Hobbs & Kehler, 1997) and (Kehler, 2002) use parallelism constraints for text coherence to identify VP antecedents. (Dalrymple, Shieber, & Pereira, 1991) and (Shieber, Pereira, & Dalrymple, 1996) apply higher-order unification to resolve the predicate variable in the semantic representation of an elided VP. (Crouch, 1999) constructs derivations in linear logic to provide alternative ways of assembling the constituents in the representation of an antecedent in order to obtain possible interpretations of the clause containing the ellipsis site.

This approach to VP ellipsis enjoys the same advantages and suffers from the same weaknesses that we noted with respect to the knowledge intensive view of pronominal anaphora resolution.

Turning to a knowledge-poor model, (Hardt, 1997) describes a procedure for identifying the antecedent of an elided VP in text that applies to the parse structures of the Penn Treebank.<sup>4</sup> It constructs a list of candidate VP's to which it applies a syntactic filter. The elements of the filtered candidate list are assigned scores on the basis of syntactic salience factors and recency.

On a blind test of 96 examples from the *Wall Street Journal* the procedure achieved a success rate of 94.8 % according to a head verb overlap criterion (the head verb of the system's selected candidate is contained in, or contains the head verb of the coder's choice of antecedent). It achieved 85.4 % for exact head verb match and 76 % for full antecedent match. A comparison procedure that relies only on recency scored 75 % for head verb overlap, 61.5 % for exact head verb match, and 14.6 % for full antecedent match.

Hardt's syntactic salience-based procedure uses essentially the same strategy and design for identifying the antecedent of an elided VP as (Lappin & Leass, 1994)'s algorithm applies to pronominal anaphora resolution. Its higher success rate may, in part, be due to the fact that recency and syntactic filtering tend to reduce the set of candidates more effectively for elided VP's than for pronouns.

As in the case of pronouns a small set of elided VP cases are not accessible to resolution by salience ranking or statistically modelled lexical preference. The following examples clearly indicate that inference based on semantic and real world knowledge appears to be inescapable for these cases.<sup>5</sup>

---

<sup>4</sup>Hard's procedure applies to elided VP's that have already been recognized. (Nielsen, 2003) presents preliminary results for the application of a variety of machine learning methods to the identification of elided VP's in text.

<sup>5</sup>(Dalrymple, 1991), (Hardt, 1993), and (Kehler, 2002) claim that the fact that inference is required to identify the antecedents of the elided VP's in 9 and 10 shows that ellipsis

- (9) Mary and Irv want to go out, but Mary can't, because her father disapproves of Irv. ((Webber, 1979))  
*Mary can't go out with Irv*
- (10) Harry used to be a great speaker, but he can't anymore, because he lost his voice.  
 ((Hardt, 1993))  
*he can't speak anymore*

## 5 The Interpretation of Fragments in Dialogue

(Fernández, Ginzburg, Gregory, & Lappin, to appear) present SHARDS, a system for interpreting non-sentential phrasal fragments in dialogue. Examples of such fragments are short answers (11), sluices (short questions, 12), and bare adjuncts (13). The latter are possible even when no wh-phrase adjunct appears in the antecedent to anchor them, as in 14.

- (11) A: Who saw Mary?  
 B: John.  
*John saw Mary.*
- (12) A: A student saw John.  
 B: Who?  
*Which student saw John?*
- (13) A: When did Mary arrive?  
 B: At 2.  
*Mary arrived at 2.*

---

resolution applies to semantic rather than syntactic representations. In fact, it is not obvious that the need for inference in (some cases of) ellipsis resolution in itself determines the nature of the representation to which the inference rules apply. (Lappin, 1996) argues that inference can apply to syntactic representations of sentences to generate structures corresponding to (i) and (ii).

- (i) Mary wants to go out with Irv.  
 (ii) Harry used to speak.

These structures supply appropriate antecedents for the syntactic reconstruction of the elided VP's in 9 and 10, respectively. The need for inference in ellipsis resolution on one hand and the nature of the level of representation to which inference and ellipsis resolution apply on the other are independent questions which should be distinguished.



- (14) A: John completed his paper.  
 B: When?  
*When did John complete his paper?*

SHARDS is a Head Driven Phrase Structure Grammar (HPSG)-based system for the resolution of fragments in dialogue. It treats the task of resolving fragment ellipsis as locating for the (target) ellipsis element a parallel (source) element in the context, and computing from contextual information a property which composes with the target to yield the resolved content. This basic view of ellipsis resolution is similar in spirit to the higher-order unification (HOU) approach of (Dalrymple et al., 1991) and (Pulman, 1997).

Two new attributes are defined within the CONTEXT feature structure: the *Maximal Question Under Discussion* (MAXQUD) and the *Salient Utterance* (SALUTT). The MAXQUD is the most salient question that needs to be answered in the course of a dialogue. The SALUTT represents a distinguished constituent of the utterance whose content is the current value of MAXQUD. In information structure terms, the SALUTT specifies a potential parallel element correlated with an element in the antecedent question or assertion. The SALUTT is the element of the MAXQUD that corresponds to the fragment phrase. By deleting the SALUTT from the MAXQUD, SHARDS produces the representation of a property from which the propositional core of the CONTENT value for the fragment can be constructed.

15 is the (simplified) typed feature structure that (Fernández et al., to appear) posit for a bare fragment phrase.

- (15) *bare-arg-ph*  $\Rightarrow$

$$\left[ \begin{array}{l} \text{STORE} \quad \{\} \\ \text{CONT} \mid \text{SOA} \mid \text{NUCL} \quad \boxed{1} \\ \\ \text{CTXT} \quad \left[ \begin{array}{l} \text{MAX-QUD} \mid \text{SOA} \mid \text{NUCL} \quad \boxed{1} \\ \text{SAL-UTT} \quad \left[ \begin{array}{l} \text{CAT} \quad \boxed{2} \\ \text{CONT} \mid \text{INDEX} \quad \boxed{3} \end{array} \right] \end{array} \right] \\ \\ \text{HD-DTRS} \quad \left[ \begin{array}{l} \text{CAT} \quad \boxed{2} \\ \text{CONT} \mid \text{INDEX} \quad \boxed{3} \end{array} \right] \end{array} \right]$$

SHARDS interprets a fragment in dialogue by computing from context (represented as a dialogue record) the values of MAXQUD and SALUTT for the assertion or question clause that the fragment expresses. It uses these feature values to specify the CONTENT feature of the clause for the fragment. The basic propositional content of the fragment clause is recovered from the MAXQUD, whose NUCL feature value is shared with the NUCL feature of the fragment clause's CONT feature.

The value of SALUTT is of type *sign*, enabling the system to encode syntactic categorial parallelism conditions, including case assignment for the fragment. The SALUTT is computed as the (sub)utterance associated with the role bearing widest quantificational scope within the MAXQUD.

SHARDS computes the possible MAXQUD's from each sentence which it processes and adds them to the list of MAXQUD candidates in the dialogue record. When a fragment phrase FP is encountered, SHARDS selects the most recent element of the MAXQUD candidate list which is compatible with FP's clausal semantic type.

16 is the Attribute Value Matrix (AVM) produced for the CONT of *Who saw Mary*.  $\boxed{1}$  is the index value of *who* and  $\boxed{2}$  of *Mary*:

$$(16) \left[ \begin{array}{l} \text{PARAMS} \\ \text{SOA} \mid \text{NUCL} \end{array} \left[ \begin{array}{l} \text{INDEX} \quad \boxed{1} \\ \text{RESTR} \quad \left\{ \textit{person-rel}(\boxed{1}) \right\} \\ \textit{saw-rel} \\ \text{see-er} \quad \boxed{1} \\ \text{seen} \quad \boxed{2} \end{array} \right] \right]$$

This is the feature structure counterpart of the  $\lambda$ -abstract  $\lambda\pi.(\dots\pi\dots)$ . The (abbreviated) AVM for the SALUTT *who* is 17.

$$(17) \left[ \begin{array}{l} \text{CAT} \quad \text{NP}[+\text{nom}] \\ \text{CONTENT} \quad \boxed{1} \\ \text{STORE} \left[ \begin{array}{l} \text{INDEX} \quad \boxed{1} \\ \text{RESTR} \quad \left\{ \textit{person}(\boxed{1}) \right\} \end{array} \right] \end{array} \right]$$

18 is the AVM produced for *John* as a short answer, where  $\boxed{1}$  is the index value of *John* and  $\boxed{2}$  of *Mary*.

$$(18) \left[ \begin{array}{l} \text{SOA} \\ \text{RESTR} \end{array} \left[ \begin{array}{l} \text{NUCL} \left[ \begin{array}{l} \textit{see-rel} \\ \text{see-er} \quad \boxed{1} \\ \text{seen} \quad \boxed{2} \end{array} \right] \\ \left\{ \textit{person-rel}(\boxed{1}) \right\} \end{array} \right] \right]$$

19 is the feature structure generated for *Who* as a sluice in response to *A student saw John*.  $\boxed{1}$  is the index value of *Who* and  $\boxed{2}$  of *John*.

$$(19) \left[ \begin{array}{l} \text{PARAMS} \left[ \begin{array}{l} \text{INDEX} \quad \boxed{1} \\ \text{RESTR} \quad \left\{ \begin{array}{l} \textit{person-rel}(\boxed{1}), \\ \textit{student-rel}(\boxed{1}) \end{array} \right\} \end{array} \right] \\ \text{SOA} \mid \text{NUCL} \left[ \begin{array}{l} \textit{see-rel} \\ \textit{see-er} \quad \boxed{1} \\ \textit{seen} \quad \boxed{2} \end{array} \right] \end{array} \right]$$

For at least some cases it is necessary to relax the requirement of strict syntactic category match between the fragment and the SALUTT to allow correspondence to be specified in terms of an equivalence class of categories.

- (20) A: What does Mary want most?  
 B: a good job/that people should like her/to have her freedom
- (21) A: When did John complete his paper?  
 B: yesterday/on Wednesday/after the teacher spoke to him

There are also instances where the scope criterion for determining the SALUTT must be overridden.

- (22) a. A: Each student will consult a supervisor.  
 B: Which one?
- b. Which supervisor will each student consult?

22b, which selects *a supervisor* as the SALUTT, is the most natural interpretation of *which one* in 22a B, even when *a supervisor* receives a narrow scope reading relative to *each student* in 22a A.

Similarly, the recency condition for selecting the MAXQUD from the list of MAXQUD candidates in the dialogue record does not always yield the correct results, as the dialogue sequence in 23 illustrates.

- (23) A: Why did Mary arrive early?  
 B: I can't tell you.  
 A: Why can't you tell me?  
 B: Okay, if you must know, to surprise you.

The fragment phrase *to surprise you* is a reply to the first question that A asks, *Why did Mary arrive early?*, rather than the second, *Why can't you tell me?*.

Knowledge-based inference is required to select the more distant question as the preferred MAXQUD in this case.

The following example for the British National Corpus is an example of a naturally occurring dialogue in which the recency criterion for determining the MAXQUD is defeasible.<sup>6</sup>

- (24) A1: That new tyre law comes in soon dunnit?  
B2: That what?  
A3: New tyre law.  
C4: First of *<pause>* first of November it came in.  
A5: Oh.  
C6: Why?  
A7: I'd better check my two back ones then.

The sluice in 24 C6 is a case of clarificatory ellipsis ((Ginzburg & Cooper, to appear)). It takes as its MAXQUD antecedent the illocutionary statement corresponding to 24 A1 rather than the statement in 24 C4.

## 6 A Sequenced Model

As we have seen work on anaphora and ellipsis within the framework of the knowledge-poor approach indicates that syntactic measures of salience combined with recency provide a highly effective procedure for antecedent identification across a wide range of ellipsis and anaphora resolution tasks in text and dialogue. These methods are computationally inexpensive and generally robust. It is possible to deal with a subset of the (significant) minority of cases which are not amenable to syntactic salience-based resolution through statistical modelling of semantic and real world knowledge as lexical preference patterns. For the remaining cases abductive inference appears to be unavoidable. These considerations suggest that a promising approach is to apply the techniques in an ascending sequence of computational cost. 25 gives the outline of a plausible architecture for such an integrated sequenced model of anaphora and ellipsis resolution.

---

<sup>6</sup>24 is from the dialogue component of the British National Corpus, File KB4, sentences 144-150. I am grateful to Raquel Fernández for providing this example.

(25)  $\langle P, Candidate\_Antecedent\_List \rangle \Rightarrow$

### **Module 1**

*Syntactic Saliency & Recency Measures +  
Syntactic & Morphological Filtering*  $\rightarrow$

Ranked Candidate List  $\rightarrow$

Confidence Metric 1  $\rightarrow$

correctly resolved;

unresolved  $\Rightarrow$

### **Module 2**

*Statistically Determined Lexical Preference Measures*  $\rightarrow$

New Ranked Candidate List  $\rightarrow$

Confidence Metric 2  $\rightarrow$

correctly resolved;

unresolved  $\Rightarrow$

### **Module 3**

*Abductive Inference*  $\Rightarrow$

resolved

The sequenced model of anaphora and ellipsis resolution proposed here moves successively from computationally inexpensive and interpretationally rough-grained procedures to increasingly costly and fine-grained methods. The model encodes a strategy of maximizing the efficiency of an anaphora (ellipsis) resolution system by invoking fine-grained techniques only when necessary.

In order to succeed this strategy must use reliable confidence metrics to evaluate the candidate rankings which the first two modules produce. Such metrics can be constructed on the model of the criteria that (Dagan et al., 1995) use to evaluate the reliability of saliency scores. When the distance between the saliency scores of the two top candidates in a list falls below a certain threshold, the ranking is taken as an unreliable basis for antecedent selection and the statistical lexical preference module is activated. Intensive experimental work using machine learning techniques will be needed

to determine optimal values for both the salience factors of Module 1 and the confidence metrics used to assess the outputs of Modules 1 and 2.

A computationally viable abductive inference component will require resource sensitive inference rules to control the size and number of the inference chains that it generates.<sup>7</sup> Resource sensitivity and upper bounds on derivations in abductive inference are essential to rendering the procedures of Module 3 tractable.

## 7 Conclusions and Future Work

While the knowledge-based and inference driven approach to anaphora and ellipsis resolution can deal with cases that require fine grained semantic interpretation and detailed real world knowledge, it does not provide the basis for developing computationally efficient, wide coverage systems. By contrast, knowledge-poor methods are inexpensive and potentially robust, but they miss an important minority of recalcitrant cases for which real world knowledge and inference are indispensable. A promising solution to this engineering problem is to construct an integrated system that orders the application of anaphora and ellipsis interpretation techniques in a sequence of modules that apply increasingly fine-grained techniques of interpretation with an attendant rise in computational cost. Confidence metrics filter the output of each module to insure that the more expensive components are invoked only when needed.

In order to implement the proposed model it is important to achieve optimization of the selection of salience parameters and their relative values through statistical analysis of experimental results. A considerable amount of work has been done on the application of salience parameters and values to minimal syntactic representations rather than fully specified parse structures. This is a fruitful area of investigation which merits further research, as it holds out the promise of major gains in efficiency and robustness for the salience methods that comprise the first module of an integrated system. Another problem worth pursuing is the generalization of lexical preference patterns to relations between semantic classes. Measuring preference in terms of semantic categories rather than specific lexical head-argument and head-adjunct patterns will increase the power and reliability of Module 2. The

---

<sup>7</sup>(Kolhase & Koler, to appear) propose resource sensitive inference rules for model generation in which the salience of a referential NP in discourse is used to compute the relative cost of applying inference rules to entities introduced by this NP. They measure salience in discourse largely in terms of the sorts of syntactic and recency factors that (Lappin & Leass, 1994) use in their anaphora resolution algorithm.

viability of the entire system depends upon determining reliable confidence metrics for both salience-based and lexical preference-based antecedent selection. Finally, to implement the the third module much work must be done to develop efficiently resource sensitive procedures for abductive inference in different domains.

## References

- Asher, N. (1993). *Reference to abstract objects in english*. Dordrecht: Kluwer.
- Asher, N., Hardt, D., & Busquets, J. (2001). Discourse parallelism, ellipsis, and ambiguity. *Journal of Semantics*, 18.
- Crouch, D. (1999). Ellipsis and glue languages. In S. Lappin & E. Benmamoun (Eds.), *Fragments: Studies in ellipsis and gapping* (pp. 32–67). New York: Oxford University Press.
- Dagan, I., Justeson, J., Lappin, S., Leass, H., & Ribak, A. (1995). Syntax and lexical statistics in anaphora resolution. *Applied Artificial Intelligence*, 9, 633–644.
- Dalrymple, M. (1991). *Against reconstruction in ellipsis*. Xerox PARC, Palo Alto, CA: unpublished ms.
- Dalrymple, M., Shieber, S., & Pereira, F. (1991). Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14, 399–452.
- Fernández, R., Ginzburg, J., Gregory, H., & Lappin, S. (to appear). Shards: Fragment resolution in dialogue. In H. Bunt & R. Muskens (Eds.), *Computing meaning* (Vol. 3). Dordrecht: Kluwer.
- Ginzburg, J., & Cooper, R. (to appear). Clarification, ellipsis, and the nature of contextual updates. *Linguistics and Philosophy*.
- Grosz, B., Joshi, A., & Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21, 203–225.
- Hardt, D. (1993). *Verb phrase ellipsis: Form, meaning, and processing*. University of Pennsylvania: unpublished Ph.D dissertation.
- Hardt, D. (1997). An empirical approach to vp ellipsis. *Computational Linguistics*, 23.

- Hobbs, J. (1978). Resolving pronoun references. *Lingua*, 44, 339–352.
- Hobbs, J., & Kehler, A. (1997). A theory of parallelism and the case of vp ellipsis. In *Proceedings of the 35th conference of the association for computational linguistics* (pp. 394–401). Madrid.
- Hobbs, J., Stickel, M., Appelt, D., & Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, 63, 69–142.
- Huang, Y. (2000). *Anaphora: A cross-linguistic study*. Oxford: Oxford University Press.
- Kehler, A. (2000). Pragmatics, chapter 18. In D. Jurafsky & J. Martin (Eds.), *Speech and language processing*. Upper Saddle River, NJ: Prentice Hall.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI.
- Kennedy, C., & Boguraev, B. (1996). Anaphora for everyone: Pronominal anaphora resolution without a parser. In *Proceedings of the 16th international conference on computational linguistics (coling'96)* (pp. 113–118). Copenhagen.
- Kolhase, M., & Koler, A. (to appear). Resource-adaptive model generation as a performance model. *Language and Computation*.
- Lappin, S. (1996). The interpretation of ellipsis. In S. Lappin (Ed.), *Handbook of contemporary semantic theory* (pp. 145–175). Oxford: Blackwell.
- Lappin, S., & Benmamoun, E. (Eds.). (1999). *Fragments: Studies in ellipsis and gapping*. New York: Oxford University Press.
- Lappin, S., & Leass, H. (1994). A syntactically based algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20, 535–561.
- Mitkov, R. (1998). Robust pronoun resolution with limited knowledge. In *Proceedings of acl'98 and coling'98* (pp. 869–875). Montreal.
- Mitkov, R. (2002). *Anaphora resolution*. London: Longman.
- Mitkov, R., Boguraev, B., & Lappin, S. (Eds.). (2001). *Computational linguistics*, 27. Special Issue on Computational Anaphora Resolution.
- Nielsen, L. (2003). *A corpus-based study of verb phrase ellipsis*. Department of Computer Science, King's College, London: unpublished ms.



- Pulman, S. (1997). Focus and higher-order unification. *Linguistics and Philosophy*, 20.
- Shieber, S., Pereira, F., & Dalrymple, M. (1996). Interactions of scope and ellipsis. *Linguistics and Philosophy*, 19, 527–552.
- Soon, W., Ng, H. T., & Lim, D. (2001). A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27, 521–544.
- Stuckardt, R. (2001). Design and enhanced evaluation of a robust anaphora resolution algorithm. *Computational Linguistics*, 27, 479–506.
- Webber, B. (1979). *A formal approach to discourse anaphora*. New York: Garland Publishing.