Chapter 7

THE LAST CHAPTER

This chapter is a miscellany. In the first three sections, I discuss some residual points and issues raised by the previous chapters. I then list some of the interesting problems that remain, and conclude with some appropriate remarks.

7.1. Anaphora in spoken language

In spoken English, vocal stress can be used to change the intended referent of an anaphor. For example, in this sentence (with normal stress) Ross gives Daryel both the measles and the mumps:

(7-1) Ross gave Daryel the measles, and then he gave him the mumps.

However, when the anaphors are stressed the meaning is reversed so that Ross gets the mumps:

(7-2) Ross gave Daryel the measles, and then HE gave HIM the mumps.

In effect the stress indicates that the referent of the anaphor is not the one you would normally choose but rather the next choice.

The principle may explain why (2-52)[1] works. If *one* ϕ were unstressed,[2] it would clearly albeit nonsensically refer to ***father***. The stress indicates that a different referent must be found, and the only place another referent can be found is "inside" the anaphoric island ***father***.

For more discussion of the relationship between anaphora and intonation, see Akmajian and Jackendoff (1970) and Akmajian (1973).

---

[1](2-52) Ross is already a father THREE TIMES OVER, but Clive hasn't even had ONE ϕ yet.

[2]Note here the interesting concept of stressing an ellipsis.

7.2. Anaphora in computer language generation

7.2.1. Introduction

Although much effort has been expended towards the understanding of natural language by computer, relatively little work has been done on the converse problem of generating a surface text from some internal meaning representation. Such generation is however necessary, for example in machine translation systems that use a language-independent intermediate representation.

Among the many unresolved issues in language generation is how best to describe an entity, and to what extent, including anaphorization, the description may be abbreviated. For example, consider (7-3) and (7-4) (based on an example from McDonald (1978b:69)), which are intended to convey the same message:

(7-3) Because of the Sangrail crisis, Ross asked Daryel to cancel his meeting with the Lesotho delegation.

(7-4) Because of the hullabaloo resulting from the theft of the Sangrail, Ross asked Daryel to cancel Ross's meeting with some people from Lesotho who had been going to inspect our taxidermy research section.

The difference between these texts is that the first is designed for an audience familiar with the people and basic issues involved, while the second is not. The first might be spoken to a co-worker, the second to a stranger met a cocktail party. In each case, different descriptions are chosen for some entities, and (7-4) avoids a pronoun which is ambiguous without knowledge of the people involved, in this case that Daryel is Ross's secretary who schedules his boss's activities.

In its most general form, description formation is an extremely difficult task, requiring the speaker to have a detailed model of the listener. In practice, so far, designers of computational systems have not used such a model, nor even given much attention to the problem; Goldman's BABEL (Goldman 1974, 1975; Schank, Goldman, Rieger and Riesbeck 1975), for example, apparently had only very primitive heuristics for description and pronominalization (though Goldman did address other important issues in the word-choice problem). Grosz (1978) and Ortony (1978) discuss some issues in generation of descriptions. To my knowledge, the only study of anaphora from the viewpoint of computational generation of language is that of David McDonald. The next sub-section is a brief description of this work.

7.2.2. Structural constraints on subsequent reference

McDonald (1978b) addresses the issue of anaphor generation, which is more constrained by syntax and sentence structure than the generation of initial reference to an entity. He describes how these constraints are used by a computer program which generates an English sentence from a tree representation based on predicate calculus. (For an overview of the program and the representation, see McDonald (1978a).)

The generation process is done in one pass without back-up. (This mirrors people's inability to unspeak the earlier words of a sentence as they generate the later ones.) When it is necessary to make reference to an element, a list of message elements mentioned so far is consulted to see if the present one has been previously referenced. If it has, a set of pronominalization heuristics are applied. First come quick checks such as whether the element has been pronominalized before. If these are unable to decide for or against pronominalization, more detailed examination takes place, and the syntactic or structural relationship between the present instance and the previous instance, such as whether they are in the same simple sentence or not, is computed.

This relationship is then used by a set of heuristics which determine whether there are any nearby "distracting references" which would cause ambiguity if pronominalization occurs. Ideally, this requires a model of the listener's knowledge; for the present, McDonald's program relies on testing the "pronominalizability" of the current element and possible distractors, and does not pronominalize if any distractor scores highest. Pronominalizability is measured simply as the weighted count of the number of pronominalization heuristics that apply to that element at that point in the text.

If an element is not rendered as a pronoun, the program must find the simplest description which will distinguish it from possible distractors. Often it is sufficient to use a definite determiner, *the* or *that*, with the head noun of a descriptive NP. See McDonald (1978b:70-71) for details.

McDonald hopes to add pragmatic and rhetorical considerations to his program. This would include using the notion of a focus or theme, pronominalization of which would usually be obligatory.

7.2.3. Conclusion

Research in anaphor generation is lagging behind that in anaphor understanding, and this is perhaps not surprising. A properly generated anaphor is one that may be quickly and easily understood, suggesting that the generator needs to consider how its audience will resolve the anaphor. It follows that the development of a proper anaphor generation system will require first the development of a full anaphor resolution system.

7.3. Well-formedness judgements

A persistent theme that has kept resurfacing throughout this thesis is the problem of knowing whether or not a sentence is well-formed. I have complained about texts alleged to prove points about the English language which are probably not English at all (see footnote 8 of Chapter 4), and about feeble attempts (my own included) to avoid this problem merely by verifying texts with a couple of readily-available informants.

It seems to me that nothing short of psychological testing is adequate to determine the relative well-formedness of a text about which there is even the slightest doubt. Language is, after all, a psychological phenomenon, and surely no-one in these modern times believes that well-formedness is a binary value engraved indelibly on a text and known to every competent speaker of the language. In fact well-formedness is a matter of degree, and no two people speak exactly the same language. It follows, therefore, that a well-formedness judgement, if meaningful at all, must represent the unbiased consensus of a number of speakers of the language.

Since the demand characteristics (Orne 1962) of informal enquiries will bias the results, it is necessary to obtain other people's judgements in a formal experiment, well controlled for influences that could bias subjects. This kind of experiment is well known in psycholinguistics; one example that we've already seen was in determining the causal valence of some verbs (see section 6.6). It is often claimed that linguistics is just a branch of psychology. Artificial intelligence is too. And both linguistics and AI need to use the experimental methods of psychology to substantiate their claims about human linguistic behaviour, upon which their theories are based.

What kind of experiment constitutes an adequate test of a sentence's well-formedness? I think that a simple speeded binary choice test would do: Subjects, told that the experiment is to determine how fast people can tell if a sentence is grammatical and meaningful, are presented with test sentences, intermixed with distractors, on a display. They have to judge the sentence and press a YES or NO key as fast as possible.[3] The proportion of subjects pushing the YES button would be a measure of each sentence's well-formedness.

You will by now be wondering if I really think that such a procedure should be carried out for each and every *John can run* sentence used as an example in the literature. After all, you object, while there are undoubtedly dubious texts for which the procedure is necessary, we highly educated and literate researchers are expert at determining what a language community, our own at least, will accept. Every time we write a sentence, whether it be an example in a linguistic argument or not, we check it for well-formedness, with almost invariable success. So why shouldn't we trust our own judgements?

My rejoinder to this is that determining the well-formedness of a text in support of a linguistic argument is not the same as determining the well-formedness of sentences used for normal communication. In the former case, one usually has the linguistic argument first and then works backward trying to

[3]This experimental procedure has been used by several researchers in psycholinguistics.

find a text which supports the point and which contains no obfuscating factors. And then, as we have seen, it is all too easy to come up with an ill-formed text without being aware of it, even if that text is as simple as, for example, (4-9)[4] Recall, too, that linguists' intuitions of well-formedness are different from those of normal people (Spencer 1973) and vary according to mood (Carroll and Bever 1978).[5] Even if the linguistic argument is inspired by an unusual real-world text, it is well to verify that this text is not unusual merely by reason of being subtly ill-formed.

I do not, of course, expect a new experimental rigour to take linguistics by storm, even though I think most people would agree with my arguments, for most linguists have neither the facilities nor the inclination to start performing experiments. A useful compromise would be a service to which linguists could send the key texts on which their arguments lie for well-formedness testing for a moderate fee.[6] [7]

> *Write a function TRANSLATE which translates the input from English to a LISP form.*
> — *Alan Keith Mackworth*[8]

7.4. Research problems

This is the traditional suggestions-for-further-research section. In it, I present some questions that remain unanswered, tasks that remain undone, exercises that the reader may find amusing. For each, the section number(s) in parentheses indicate where in this thesis the matter is discussed further.

The study of language and reference:

- (1.1) Is an implementation a theory?
- (1.2) How do words denote concepts?
- (1.1) Can we define a (domain-independent) Habitable English for database queries? (Habitable English is to grammar, semantics and pragmatics as Basic English is to vocabulary.) Is there a simple formula, similar to those

[4](4-9) John left the window and drank the wine on the table. It was brown and round.

[5]Moreover, I have occasionally been surprised by the poor linguistic abilities and/or minimal communicative competence of some of AI's "amateur linguists".

[6]World-wide franchises are now available. Contact the author for details.

[7]Nothing in this section is to be construed as belittling the important theoretical aspects of linguistics. One reader of a draft of this section suggested that just as experimental physics needs theoretical physics, so linguistics needs the important insights gained from theoretical work which cannot be supplanted by any amount of experiment. This is true. However theoretical physics has its theories tested by experimental physics. My complaint is that linguistic theories are often accepted without any attempt at experimental verification, and this is a Bad Thing.

[8]Part of an assignment for third-year UBC Computer Science students learning LISP, 17 November 1978.

used to determine the readability of a text, which could measure habitability without recourse to performing real-world experiments with the language subset?

- (3.2.7) Write a book discussing issues in the relationship between the nature of language generation and understanding, and the structure of the human mind.
- (4.2) How do oenologists communicate?
- (5.6) Can natural language be understood by a system using a finite set of rules, or a finite set of rules for generating a possibly infinite set of rules?
- (7.3) Write a critique of my remarks on the need to psychologically test the well-formedness of sample texts, presenting an opposing view.
- (7.3) Buy a sample text testing service franchise from the author, and see if it proves to be useful and/or profitable. Has your service influenced linguists' attitudes to sample texts?

Anaphora, anaphors and antecedents:

- (2.1) Can the set of implicit antecedents that texts can evoke be formally defined? What may be an implicit antecedent, and under what circumstances? Consider especially antecedents for verb phrase ellipsis.
- (2.3.1) Formalize the conditions under which *same* can be used as an anaphor.
- (2.3.2) Formalize rules for the generation and analysis of surface count anaphors.
- (2.3.7) Come up with an elegant theory explaining all usages of the non-referential *it*. Explain why sentence (iv) of footnote 38 of Chapter 2 seems ill-formed.
- (2.4.2, 6.7) What non-inferred reference relations are possible? What is to be done about semantic distance?
- (2.6, 6.5, 6.6) Investigate default antecedents. Are they affected by any factors other than plausibility and theme? How do they relate to verb causality?
- (6.4) Formalize rules for syntactic and semantic parallelism.
- (6.5) How can plausibility of a candidate antecedent be efficiently measured computationally?
- (6.6) Are causal valence data of any computational use? Can the concept of causal valence be usefully generalized?
- (7.1) In what ways can stress on an ellipsis be phonetically realized?

Anaphora resolution systems:

- (3.1.6) How may an anaphor resolver best be evaluated? Prepare a standard corpus of text, which includes all types of anaphora and reference both easy and hard, and make it available to people who want independent test data for their theories or systems.
- (3.1.6) Beef up Hobbs's algorithm so that it works even more frequently.

- (3.2.3) Can an anaphor resolver do without heuristics?

Focus and discourse theme:

- (3.2.1) Should there be one large focus set, or should focus be divided up into noun types, verb types, etc? What is the best such division?
- (4.1 passim) Define the concepts of theme, rheme, topic, comment, given and new so definitively that everyone will use your definitions.
- (4.1 passim) How can the local and global theme of an arbitrary text be determined computationally?
- (4.2, 5 passim) What exactly IS the relationship between theme and focus?
- (6) To what extent should a focus be computed independent of any anaphor that needs resolution?

Current approaches to anaphora and focus:

- (5.1) Generalize the concept of secondary competence. Is there any psycholinguistic evidence that linguistic competence and/or verbal ability comes in well-defined layers? Are some people more prone to generating inconsiderate anaphors than others? Do such people actually find inconsiderate pronouns easier to understand than other people do? Could there be a consistently different model of language in such people?
- (5.1.1, 5.1.2, 7.3) Test Kantor's assertions about pronoun comprehension through experiments such as observation of readers' eye movements and/or reaction-time measurement.
- (5.1.1) What factors affect the activatedness of a concept?
- (5.1.2) How do we know when a concept occurs only as a descriptor and not "in its own right"?
- (5.2.1) Are there other common sorts of dialogue which are as highly structured as task-oriented dialogues? How can their structures be exploited?
- (5.2.2) How could Grosz's methods be applied to the resolution of pronouns?
- (5.2.2, 5.3.3) Given a sentence in a vacuous context which sets up a theme or focus for the interpretation of subsequent sentences, how may this theme be discovered? That is, how is an initial focus determined?
- (5.2.2, 5.3.3) Analyze and classify various clues to focus shift, and give rules for their detection. If more than one is indicated, how is the conflict resolved?
- (5.1.2, 5.2.3, 5.5) Can Grosz's mechanisms be generalized?
- (5.2.1, 5.3) Is focus, the repository of antecedents, really identical to the focus of attention or the discourse topic? If not, under what conditions are they identical?
- (5.3.2) How can a language understander decide when a difficult reference can be left unresolved without engendering problems later on?
- (5.3.4) What is the relation between the genericity of an anaphor and its antecedent?

- (5.4.1) Formalize a complete solution to the intrasentential anaphor resolution problem in Webber's formalism.
- (5.4.2) How may a *one*-anaphor be reliably recognized?
- (5.4.2) Are all antecedents of *one*-anaphors textually recent? Under what conditions are textually recent descriptions not available as antecedents?
- (5.4.2) Find the general principle by which strained anaphors can be resolved.
- (5.4.2) Under what conditions can list elements be abstracted into an antecedent for a *one*-anaphor?
- (5.4.3) How may inference be used with Webber's formalism so that verb phrase ellipsis triggers that are not textually similar to the elided VP may be detected?
- (5.4.4) To what extent does Webber's formalism need the addition of discourse pragmatics? How could they be provided?
- (5.5) Can scripts or frames be made suitable for the understanding of free or deviant discourse?
- (5.5.1) What is the "right" set of discourse coherence relations *(a)* for anaphor resolution, and *(b)* for general NLU? Define them rigorously.
- (5.5.1) Can a set of primitive coherence relations for building more complex relations be defined? Be sure to give the rules under which the primitives may combine.
- (5.5.1) What is the best level – clause, sentence or paragraph – to handle discourse cohesion?
- (5.5.3) Is the search order for a node for feasible connection in Lockman's (1978) CRRA always optimal? Can it lead to error?
- (5.5.3) Can Lockman's CRRA be sure all referable entities are considered?
- (5.5.3) Can the sub-tree of a complex sentence always be determined syntactically? Look for counterexamples to Lockman's table look-up procedure.
- (5.5.3) Devise and implement a judgement mechanism for Lockman's CRRA.
- (5.6.1) How can the temporal location of a text be determined?
- (5.6.1) Under what conditions can a tenseless text contain temporal anaphors?
- (5.6.2) Is there a natural language that has a locative equivalent to tense? (May require field work.)
- (5.6.2, 5.6.1) Is the *now* location of a text ever an obfuscating factor as the *here* location sometimes is?

Anaphora in discourse generation:

- (2.2, 7.2) What sort of model of the listener does a speaker have to have for anaphor generation? What knowledge representation is appropriate for the model? Does the model have psychological reality? How does the model relate to Cohen's (1978) work on models of discourse participants?

7.4 Research problems

- (7.2) Should a discourse generator operate in one pass without back-up?
- (4.1 passim, 7.2) Devise a generative grammar in which local and global theme are explicit elements in the deep representation. Use your model to construct a computational discourse generation program for a machine translation system.
- (7.2) Devise a mechanism which uses an audience model in generating descriptions and anaphors in discourse. Integrate it into the program you constructed in the preceding exercise.

7.5. Conclusion

This thesis has surveyed the problem of computational understanding of anaphora and attempts at a solution thereof. We have seen that an adequate solution to the problem will require the use of discourse pragmatics and the notion of theme to maintain a focus. We have further seen that a complete solution, in which all reference relations, including those determined by inference, are recovered is extremely difficult, and the surface has yet barely been scratched. The work that remains to be done will influence and be influenced by work in linguistics and artificial intelligence. Anaphora buffs have an exciting time ahead.

> *English has no anaphors and the whole notion of anaphora has simply been a popular fallacy.*
>
> — *William C Watt (1973:469)*