Visiting ICOT: A Trip Report

Manny Rayner, SICS

1. Introduction

My visit to ICOT was arranged at short notice, as part of a reciprocal agreement with SICS concerning exchange of visiting researchers, and was also very brief (only six weeks). It thus had a fairly informal nature, and as far as I was concerned essentially the flavour of an initial exchange of information between ICOT Second Laboratory and the Knowledge-Based Systems group at SICS. If I am to summarize in a couple of sentences what the positive results of my trip have been, I think it fair to say that each of these organizations now has a much better idea of the research presently going on at the other one, and the way in which it interacts with their own work. I very much hope that this will lay the foundations for future co-operation between ICOT and SICS.

Since my main speciality is Natural Language Processing (NLP), the greater part of my activity here naturally fell under this heading. I was given a detailed presentation of the DUALS-III and CIL projects by several of their members, in particular Dr. Sugimura, Dr. Hasida, Dr. Mukai, Sano-san and Hatano-san. I myself in two seminars presented the work I have done at SICS, and especially the following topics: the SNACK-85 question-answering system, the use of explanation-based generalization (EBG) in NLP, and the semantics of comparative constructions. The last of these in particular I have discussed at great length with Dr. Hasida, and I describe the results of this more fully in Section 2. I also carried out some experiments on large-scale application of EBG to NLP, which are described in section 3.

Outside NLP, I gave a seminar on the work I have performed on temporal reasoning, which has mainly had the character of a series of refinements and attacks on the "Yale School", typified by Drew McDermott and Yoav Shoham. I got some quite interesting feedback at this lecture, in particular from Dr. Mukai. I was also given the opportunity to examine in detail the experimental I-Go system developed by Dr. Sei. Later, I had a short but extremely interesting discussion, also with Dr. Mukai, about the application of sheaf-theory to the modelling of dynamic systems; this turned up the surprising possibility of a connection with Dr. Sei's work. I attempt to summarize my side of the story in section 4.

In addition, I was given a demonstration of the CAP and CAL systems, this being not so much for my own benefit as for that of my colleagues Lars-Henrik Eriksson and Dr. Philippe Mathieu, who are now in contact with the relevant people at ICOT. From what I have understood them to say, this may possibly lead to some kind of fruitful interaction in the future, though this really lies outside my area of competence.

2. The Syntax and Semantics of Comparatives

This subject has interested me for some time now, and together with Amelie Banks (formerly of the University of Uppsala; now employed at ZYX Sweden AB, Stockholm), I have written three papers on it. The first two were presented at the Vancouver Workshop on Logic Programming and Natural Language Understanding (1987) and the Buffalo ACL conference (1988); the third (whose content properly contains that of the earlier ones), has been issued as the SICS report *An Implementable Semantics for Comparative Constructions*, and is currently under review by the international journal *Compuational Linguistics*. It formed the basis for the discussions with Dr. Hasida mentioned above.

Essentially, the main thrust of the argument put forward by Ms. Banks and myself is that (contrary to the precepts of Transformational Grammar), comparative constructions are generally not elliptical in nature, but can be interpreted directly in a monostratal theory. The thesis is not novel, and has previously been argued for in certain cases by, among others, Hankamer, Postal and Pinkham; however, (as far as I am aware) we are the first people to present the idea in a systematic form, and use it as the basis of a functioning, implemented program. What Dr. Hasida, Sano-san and I have discussed is the extent to which these ideas are applicable to work being done at ICOT: this in practice means adapting them to a Situation-Semantic/JPSG treatment of Japanese.

Limited time, and my own lack of experience with Japanese and JPSG, has unfortunately not permitted us to give this question a definite answer as of the present moment. The current state of affairs is as follows: Dr. Hasida has written a working paper, in which he presents in outline the results of our discussions, supplemented by a sketch of their realization in JPSG; and I have implemented a very small question-answering system for Japanese (loosely based on the one Banks and I developed for Swedish), which can handle a few of the simpler comparative constructions. The files for this can be found on the subdirectory `manny/comparative` on `icot21`. Our current plan is to continue to correspond by E-mail after I return to Sweden, and aim to finish a joint paper in time for the December 1st submission deadline for COLING-90. In view of the progress already made, I do not expect this to cause undue difficulty.

3. Explanation-Based Generalization and NLP

The technique of explanation-based gneralization (EBG) has recently become the focus of considerable attention in Machine Learning circles. In a paper presented at the FGCS conference last December, I suggested that NLP ought to be a promising application area for EBG. I illustrated my point with examples using logic grammars which, though small, were not by any means trivial in size; none the less, it was fairly clear from the referees' comments that something larger would be necessary if people were to be convinced of the general utility of the scheme. In particular there were three main questions to be answered:

1) EBG as I deal with it is only applicable to pure Horn-clause programs. Can NLP systems large enough to be taken seriously be "cleaned up" (i.e. converted to pure

Horn-clause form), without either catastrophically lowering performance or requiring the rewriting of large sections of code?

2) Even if 1. above can be answered affirmatively, can the extraction of learned operators be performed with reasonable expenditure of time and space?

3) As S. Minton pointed out last year in a much-remarked-on paper from the AAAI-88 conference, it is very important with EBG to establish that the derived operators are actually worth the trouble of acquiring. If the average amount of time lost in unsuccessful look-ups of an operator exceeds the average saving when the operator can be applied, the operator is no use. What is the situation here with respect to NLP?

The experiments I have performed at ICOT have involved the application of an improved version of the EBG interpreter from my FGCS paper to Fernando Pereira's well-known CHAT-80 system. Although CHAT-80 is of course still quite small compared to the largest NLP systems, it is definitely a non-trivial program; my feeling is that it is large enough to constitute a serious test. My results (which I plan to write up as soon as possible, hopefully for the Prolog/NLP workshop at SICS next month), are as follows:

1. "Cleaning up" CHAT-80, by removing cuts, was a simple business which only required a day or two of routine hacking; in all cases, the only changes needed were some minor adjustments generally involving a modest increase in the size of the code. The effects on performance were not serious. In other words, it seems that NLP systems are inherently fairly clean; not a surprising result, of course, given the current interest in unification-based grammar formalisms.

2. The EBG interpreter runs easily fast enough to make derivation of rules from large example sentences practical. This, again, is not very surprising; since it is essentially a normal Prolog meta-interpreter, the difference in speed ought to be only a reasonably small constant.

3. This question can unfortunately not be answered properly without a large amount of statistical data concerning frequency of different syntactic patterns. However, my initial results are promising. Since many "function-words" like "a", "the", "and" and "more" are unique (do not generalize), the learned rules tend to be such that incorrect application are in most cases detected by simple matching, and thus in very little time. It is also possible (though I have not yet implemented this), to arrange the rules in a TRIE structure keyed off non-generalized words and part-of-speech information. This ought to give search-times logarithmic in the number of rules, which in practice seems to give the possibility of learning at least thousands of rules before reaching the point of diminishing returns. But even without this refinement (i.e. with sequential search, as currently implemented), it would appear that more than two hundred rules could be acquired without swamping the system.

4. The Application of Sheaf-Theory to Located Games

I will start this section by stressing that all the ideas presented are extremely preliminary in nature, and have by no means been worked out properly. However, since Dr. Mukai was kind enough to take them seriously, I am happy to use this

opportunity to sketch out my thoughts on the subject.

In a paper from 1986, F. Pereira and L. Monteiro suggested that sheaf-theory could form a good basis for the study of concurrent systems. The fundamental notion was that a family of interacting concurrent processes could be regarded as the underlying set of a sheaf; the idea was that the behaviour of a sub-process could be taken as the *restriction* (in the sheaf-theoretic sense) of the behaviour of a super-process. Interaction between processes occurs due the presence of shared sub-processes.

My suggestion is that this could be a good way of thinking about the game of I-Go. Despite fairly considerable efforts, there is still no I-Go program of even weak amateur standard; I easily defeated Dr. Sei's program, probably the strongest in the world, on a maximum handicap of nine stones. Since I myself am far from being a strong player, this would seem to indicate that some basic method is lacking in the implementation.

Study of textbooks written by strong I-Go players reveals that the concept of *aji* (potential sequence) is fundamental to the human understanding of the game. Dr. Sei assures me, however, that his program so far has no representation of aji. I suggest formalisation of aji as a tree of moves occurring under certain conditions in some local part of the board, and thus in the sense of Pereira and Monteiro associated with it. The point is that it is often highly desirable to consider the restriction of this tree to some sub-area: thus although the tree itself may be complex, the restriction can be simple. For example, a very useful concept is that of a *sente* (forcing) move. This can be thought of as follows: we have an area P, and a piece of aji A in P. P' is a subset of P, and the restriction of A to P' is the tree (to invent some hopefully fairly transparent notation) :

```
[Black plays at x,
 (either: White plays somewhere in P-P' or
          [White plays outside P,
           Black plays in P and makes a gain])]
```

We say that x is a sente move against P. (Obviously, this is still very schematic). The point is that if P' is the intersection of P with some other area where things are happening, the restriction can be much more useful than the whole tree, since it summarizes what we want to know; as long as nothing has happened in P yet, Black's playing x will give White the choice between accepting a loss in P or giving Black an extra move in the area outside P.

Another example of how this kind of reasoning can work is the process of establishing that a group has been encircled. Normally, encirclement is not absolute, in the sense that the encircling stones form an unbroken string; rather, there are a number of gaps, none of which, however, can be broken. To formalize this, we can consider associating a piece of aji with each potential break-through area, encoding how each specific break-through attempt can be parried. If these areas intersect, it is necessary to show that the restrictions of the different pieces of aji are consistent on the intersections. This, I hope, should be much easier than doing a global analysis of the situation, and also be close to what a human player does.

I understand that Dr. Mukai may be visiting Sweden during the early part of the

summer. We have agreed in this case to discuss further the ideas mentioned above.

5. Conclusion

Both I and my family have enjoyed our stay in Tokyo very much, and we wish to thank all the people who have displayed us such kindness in so many different ways. It seems almost impossible to list them all here, and for fear of offending someone by omitting their name I shall refrain from attempting to do so. I can only say that these six weeks have been one of the memorable experiences of our lives, and that they were over all too soon; we are already trying to think of some way to arrange another visit to Japan! Before that, however, I sincerely hope that we will be given the opportunity to return some fraction of your hospitality on our home ground.

Curriculum Vitae

Manny Rayner

Born 11th June 1958 (London, England)

B.A. (Mathematics) from Cambridge, England, 1978

Researcher, UPMAIL, Uppsala, Sweden, February 1982 - February 1986

Researcher, SICS, Stockholm, February 1986 - present

Currently completing Ph.D. at SICS.

8 published papers (I sent a list by facsimile)

Research interests: Natural language (especially semantics), logic programming, temporal and modal logic, planning, machine learning.