Report on a Visit to ICOT

Jean Mark Gawron
CSLI, HP-Labs

1 Introduction

I was first invited to ICOT two years ago after work on Situation Theory and Situation Semantics. At that time I was forced to decline for personal reasons. It was a genuine pleasure to receive another invitation to come from Hasida-san, after I had spoken to Harada-san of Waseda University about my interest in a visit. I accepted Hasida-san's invitation, and after one date fell through because of illness, I finally arrived at ICOT on March 18.

My work in Situation Semantics has centered on inssues connected with situatedness of language and the development of a general framework for compositional semantics which accounts for that situatedness. In the last few years I have been especially interested in the phenomenon of anaphora. I have also worked for a number of years within the linguistic framework of HPSG, helping to build and refine the system implemented at HP Labs. There, too, my central interest has been semantics, especially, in the last year the semantics of comparatives.

One of my goals in visiting ICOT was to see whether researchers with such a wide range of interests were converging on a view of NL-processing similar to that held by myself and my colleagues at HP labs. In brief, that view is that the solution of some central problems facing researchers in Natural Language is not specific to NL, but involves general techniques of inference and constraint resolution.

As I see it, the outstanding problem in the Natural Language Processing is the integration of all the information of very different kinds which goes into the understanding of an utterance, from a classification of the sounds into linguistically meaningful units to the identification of a larger rhetorical goal in the service of which that utterance was made. Linguistic "pipeline" models encourage the fiction that all this information is processed sequentially up a kind of linguistic chain of being, from the phonetic through the semantic and pragmatic, with each step dependent on the sucessful completion of the previous one. Yet I think it's safe to say that very few if any of those in the Natural Language processing community believe that this model is a serious one – either for modelling human processing or, simply, as a practical matter, for dealing with the complex information flow of real communication situations. What is going on when Natural Language understanding happens clearly shows the interaction of a network of constraints of very different kinds, not in any predefined sequence but in a variety of dynamically changing ways. Let me call problems which have this character Heterogeneous Constraint Problems (HCP's). Any system which purports to deal with the information flow of Natural Language will have to be able to solve HCP's.

2 Summary of Visit

The morning of my first day at ICOT Iwata-san gave me a very helpful overview of all the research going on here at ICOT. I got some idea of the breadth of the topics covered here. Although it bears no obvious relation to mty chosen field of study, I was especially taken with the idea that the mapping of the human genome offers a promising application for the parallel architectures and machines under development here.

I gave three talks in all at ICOT, all on the first three days of my visit, because there were meetings of PSG and STS going on then. The first talk was entitled, "Three Approaches to the Interpretation of Anaphora," the second "E-Type Anaphora in Situation Semantics," the third, "Comparatives in HPSG." During my visit I had discussions with a number of ICOT researchers, most of whom are referred to below. During the first three days when the meetings were going on, however, I had some very fruitful discussions with some non-ICOT researchers, especially Harada-san of Waseda University and Kawamori-san of NTT, while at the meeting and sightseeing.

The remainder of my time at ICOT was taken up with discussions, demoes, and the writing of this report. My interests naturally led me to focus on the Third, Fourth, and Sixth Laboratories. I had demoes from Terasaki-san and Aiba-san, Satoh-san, Hasida-san, Tsuda-san, and Fukushima-san. I had a particularly fruitful series of exchanges with my host Hasida-san.

But useful exchanges were not limited to my area of expertise. Hearing of my amateur interest in biology Tanaka-san of the Third Laboratory found me some papers on the formal language properties of nucleotide chains (they seem to require index grammars). He also told me about the Protein Function database being written in Quixote.

3 Constraint Logic Programming

3.1 CLP

As I said in the introduction one of my chief interests in coming to ICOT was to see what aspects of the work here were contributing to the solution of HCP's (Hetergeneous Constraint Problems), because I believe NL understanding is a paradigm case of such a problem. HCP's have two independent features:

1. they involve the simultaneous solution of complex constraints of different kinds (for linguists, semantic, pragmatic, syntactic, phonological)

2. There is in general no sequence that can be assigned to these constraints. They interact in context-dependent ways; the flow of information is multi-directional.

Solving problems with these properties (especially property 2) makes it all the more important to effect a separation which has long been one of the priorities of logic programming: the control structure which solves constraints should be kept separate from the constraints themselves. The idea here is twofold: first, the best defense against complexity is clarity, and this separation makes things clearer; second, the same control structure is supposed to be general enough to deal with numerous different constraint-types. Obviously, in the face of property (2), it becomes crucial to have a single control structure that can accomodate information from different kinds of constraints.

Thus, I was especially interested to see what work at ICOT was pursueing this goal of Logic Programming, and trying to circumvent some of the limitations of Prolog in this respect. The demo Aiba-san and Terasaki-san gave of CLP gave me a starting point. In CLP the general control structure is a first-order equation solver, which will output a set of solutions in canonical form, some of which may themselves simply be constraints on variable bindings. The attractive feature here is a single framework for attaining both complete and incomplete solutions. When we consider the total range of possible effects sentences can have, "total" solutions of the understanding problem may be much more the exception than the rule.

3.2 HCLP (Soft Constraints)

I was treated to a demonstration of CHAL by Satoh-san. Here the general goal was one that has much relevance to the Natural Language processing, to incorporate into the constraint language framework the idea of "soft" constraints, ordered constraints that can be relaxed when necessary to satisfy the constraint-set as a whole.

First, I think this is very much the right approach towards acheiving non-monotonic effects in inferencing. In other words, an approach that uses classical logic together with a sophisticated view of what is going on when defeasible conclusions are reached is more attractive than one which abandons classical logic in favor of something with much shakier foundations. The "sophisticated view" in this case is the idea that defeasible conclusions are "best" solutions with respect to the current constraint set, with all hard constraints satisfied, and as few soft constraints as possible violated.

Two directions suggest themselves for developing this research. First, work on system-building principles seems crucial. The task of writing complicated constraint sets (such as Natural Language grammars) has already proved itself very difficult. We now add to the problem that of choosing rankings among soft constraints. Two sorts of solutions to analogous complexity problems have been offered in the literature:

- Enunciate principles which allow us to determine of a constraint in *isolation* what its ranking properties are.
- Design systems that can learn soft constraints and assign appropriate rankings by exposure to data. (One of the great appeals of Connectionist models is that they do this.)

Satoh-san's paper already announces his interest in the second kind of solution. The first kind may be equally important. Of course, the two kinds of solution are not mutually exclusive.

Second, generalizations of the hierarchy assumptions in this paper. It seems to me unlikely that the combined constraints for Natural Language processing can be partially ordered once and for all. Rather it seems more likely that various kinds of CONDITIONAL ordering will be necessary, which suggests that constraint satisfaction algorithms will have to be much more sensitive to the dynamic informational state. In state S constraint one may outrank constraint two. In state T, the reverse obtains. Such conditions can of course be accomodated into the CHAL framework by conditionalizing (and therefore complicating) the constraints themselves. If C1 outranks C2 when S and C2 outranks C1 when T then in CHAL we might replace C1 and C2 with :

(C1'): If S then C1
(C1"): If T then C1
(C2'): If S then C2
(C2"): If T then C2

We can then impose the ordering:

C1 ' outranks C2 '
C2 " outranks C1 "

But we would like to be able to do this without complicating the constraints and without, for example, having to verify S every time we use (C1') (State S might just be defined by this very ordering of C1 and C2). Moreover, ordering conditions may involve other constraints. When C1 and C2 hold, for example, C3 outranks C4.

Concerns like this last one are why I am very interested in work that not only deals with relaxable (soft) constraints, but also tries to provide a framework that allows dynamic relationships among them.

4 cu-Prolog and Dependency Propagation

4.1 cu-Prolog

Tsuda-san gave me an interesting demo of cu-Prolog and JPSG parsing. Here the issue is a generalization of Prolog that allows a less procedural handling of constraints. I am not competent to assess the interest of the relationship between cu-Prolog and ordinary Prolog. The interesting feature to me was the possibility of representing some ambiguity as a parse tree together with a constraint with multiple solutions. My understanding is that the cu-Prolog version of JPSG can do this for feature-structure ambiguity, anaphoric binding ambiguity (the binding of *zibun*, for example), and lexical semantic ambiguity.

What this immediately points to is the desirability of a framework in which all ambiguity could be dealt with in this way: a certain amount of structure attached to constraints which encode the consequences of disambiguating that structure. It is in taking steps in this direction that the interest in DP lies.

4.2 Dependency Propagation

As mentioned in 3.1, a paradigm goal of Logic Programming is to separate control structure from constraints. The work in DP appears to be an effort to push this trend as far as it can go. This work is very much a return to foundations. Everything that once looked easy now looks very hard. But this should not be particularly surprising. I once heard Marvin Minsky give a talk in which he explained that the general direction of AI is rather the reverse of what one might think. Early work in AI focused on solving problems in calculus, and later algebra; it was only when a certain amount of sophistication had been attained that a program that played with blocks could be written.

First the notion of constraint is extended from Horn-Clauses to any first order constraint; with this move, general decidability is given up; what is gained is expressive power. To the general question of *what do I do next* that is posed for a general purpose constraint solver, DP proposes the answer *resolve dependencies,* where dependecies are terms linked by sequences of equalities. A general principle of truth-maintenance is thus appealed to. To further control choices among various dependency resolutions, DP proposes certain heuristics, some of which at least, are domain-independent. Finally, finer-grained choices will be made on the basis of certain cost-assignments to atomic constraints, which are interpreted as measures of potential energy, along with a general principle: minimize potential energy.

The enterprise is quite ambitious, but the potential rewards are striking. Most importantly, from my point of view, this program seems to be directly geared to dealing with HCP's: the theory of potential energy allows for dynamically altered priorities among constraints. Yet the general control structure will still allow information flow between coinstraints of very different kinds.

The papers which I have read, illustrated with some simple parsing and generation problems are quite suggestive and impressive. After a number of conversations with Hasida-san, I am convinced that this is a very promising line of research, with great potential for Natural Language applications, and for other HCP's as well. I have some thoughts and speculations about the enterprise.

First, an observation. It seems to me that the framework introduces in addition to constraints something we might call meta-constraints (constraints about constraints). The costs assigned to atomic clauses are one sort of meta-constraint. The disjunctive links between clauses are another. The former may change dynamically, the latter may not. Part of the task of writing a constraint set must now include specifying meta-constraints. Some of the comments made in 3.2 about the potential complexity of ordering CHAL constraints may also be made about the problem of assigning meta-constraints in DP.

Second, a remark about the theory of Cost Assignment. One slogan that is used in the programmatic paper by Hasida-san and Matsubara-san is that heuristics for what sorts of computation should be carried out are really the same heuristics for which sorts of things are most plausible. I don't think this is true in general, because ease of verification is independent of plausibility. Perhaps a simple example will help illustrate the point. Suppose we are in a situation where we know the exclusive disjunction **a** XOR **b** XOR **c**. We have ways to verify **a** and **b**, but no way to directly verify **c** (except by disjunctive syllogism). Our task now is to determine which of **a** or **b** or **c** is true. Now suppose the probability of **a** is .8 but the computational cost of veryifying **a** is 100. Suppose that for **b** the probability is only .1 but the computaional cost is a mere 1. Then we are faced with the problem of deciding between two strategies in deciding among **a**, **b**, and **c**: (A) verify **a** first or (B) verify **b** first. If we go purely by plausibility we chose the (A) strategy. But incorporating ease of verification into our calculation yields a different result. The average cost of decision for the B strategy will be 91 ((.1 x 1) + (.9 x 101)), and 100.2 for the A strategy. So by this average cost criterion, verifying the less plausible aleternative first is favored. If this notion of cost makes sense in the DP framework, then that ought to have some consequences for how costs are combined (probabilities will combine mulitiplicatively, computation cost, at worst, additively).

Third, the radicalness of the framework raises some fascinating foundational questions. The most basic is this: I am no longer sure what counts as equivelent sets of constraints. To be sure, we still have the same model theoretic apparatus as before, so that constraints true in all the same models may be called equivalent. But meta-information about constraints may yield quite different behaviors for constraints sets that are equivalent in this sense. Indeed, in his paper, "Potential Energy of Linguistic Constraints", Hasida-san raises the interesting possibility of relieving linguistic constraints of some of their burden (he mentions the Subcat principle of HPSG) by transferring some constraints into meta-constraints or general heuristics for the control mechanism. This means we had better have a notion of performative equivalence (equivalent behavior in equivalent informational states) quite independent of (and perhaps more relevant than) our notion of model theoretic equivalencve.

5 Natural Language Generation

Fukushima-san gave me a demo of a Natural Language generation system that constructed logical arguments on the basis of facts stored in a Quixote knowledge base. The interesting feature of this system to me was the level preceding semantic representation, in which rhetorical roles like thesis, antithesis, and reason were represented. These are discourse, not logical relations, which emerge from a plan generation component: the thesis/reason relation, for example, is not one of logical consequence but one that holds between two propositions when one provides evidence for the other. Thus, it will accomodate defeasible conclusions. This linkage between plan generation and discourse function seems to me to be in line with the most interesting work in generation.

6 Parallel Processing

On the last day of my visit I saw the videotape on the Multi-PSI system. I am not competent to judge the particulars of this parallel machine as against others, but I remain convinced that the general program of coupling research on inferencing with research on parallel computation is an intelligent one. In particular, the general logic programming goal of decoupling constraints from control structure seems to provide a good starting point determining the degree of sequentiality in a program, and thus the degree to which Parallelism will help. My general impression, for what it's worth, is that if technical problems involving intense communication between processors can be alleviated, then parallelism helps considerably for problems like NL-parsing. Whether other kinds of constraint resolution in NL -understanding will benefit equally doubtless depends on the degree of sequentiality in those problems. These seem to me to be very open issues.

7 Conclusion

There is no real conclusion to be reached about my visit to ICOT. A number of different discussions happened; I saw too much for me to understand all at once; and I will spend a lot of time in the future thinking over some of the questions that were raised here.

Yet if I had to name one domain in which cooperation among Japanese and American researchers in AI and Natural Language Processing could be most fruitful, I would say something like this: the outstanding problem in NL research right now is that NL presents what I called in the introduction an HCP.

It seems to me that a number of different lines of research being pursued at ICOT may help contribute towards the solution of HCP's. Research into constraint programming languages with soft constraints as in CHAL is fundamental. Beyond that, the kind of generality and flexibility sought in DP will, I think, be necessary to handle the full complexity of ambiguity resolution.

8 Thanks

This is the place to thank ICOT for inviting me and giving me the opportunity to engage in the stimulating exchanges of the last two all-too-short weeks. My thanks to Fuchi-san for his kindness as a host, and for a wonderful welcoming lunch, and to Yokota-san, and Tanaka-san, whose laboratories I spent most of time in, and who presided over an unforgettable farewell dinner of eel. Thanks also to Hasegawa-san, whose researchers in the Fourth Lab showed me CLP and CHAL. Special thanks go to my thoughtful host Hasida-san, with whom I had many discussions that will provide food for future thought, and whose knowledge of Tokyo and Tokyo-eating was invaluable.

I look forward to further fruitful exchanges with the people now working at ICOT, both on visits to Japan and in the States.

JEAN MARK GAWRON
Curriculum Vitae

Center for the Study of Language and Information
Stanford University, Ventura Hall
Stanford, CA 94305

EDUCATION

Ph.D. Linguistics. University of California at Berkeley. December, 1983.

M.A. Linguistics. University of California at Berkeley. June, 1980.

B.A. French. New York University. June, 1974.

DISSERTATION TITLE

"Lexical Representations and the Semantics of Complementation"

LANGUAGES

Polish, French, Italian

COMPUTER LANGUAGES

Lisp, some Prolog

FELLOWSHIPS AND AWARDS

FLAS in Polish, Fall, 1980, Spring 1981; Fall, 1979 - Spring 1980.

Bella Zellerbach-Cross Scholarship and Maria Gudde Scholarship, University Fellowship, Fall 1977 -Spring 1978.

RESEARCH EXPERIENCE

Visiting Researcher, Deutsches Forschungszentrum fur Kunstliche Intelligenz. Universitat des Saarlandes. September, 1990 - 1990.

Post-Doctoral Fellow, Center for the Study of Language and Information, Stanford University. August, 1985 - 1989. Projects: Situation Theory and Situation Semantics, Lexicon Initiative.

Consultant HPSG Natural Language Project, Hewlett-Packard Laboratories, 1501 Page Mill Road, March 1987 - present.

Associate Research Scientist, Courant Institute of Mathematical Science, New York University. 1984-85. DARPA Strategic Computing Project on the Processing of Military messages. PI: Ralph Grishman. Wrote chart parser, worked on semantic procesor, and wrote grammar with treatment of coordination, long-distance dependencies.

Leverhulme Fellow, University of Edinburgh, Department of Artificial Intelligence. 1983-84. Designed semantic processor and semantics for the Edinburgh GPSG system.

Member of Technical Staff, Computer Research Center, Hewlett-Packard. GPSG Natural Language System. June, 1981 - October, 1983. Worked on all modules of Database Interface System, including grammar and lexicon, semantic processing, and knowledge representation.

Research Assistant, Electronics Reasearch Laboratory, Department of Computer Science, UCB, NSF grant to study Fuzzy Logic. Fall, 1978 - Winter, 1979. Principal Investigator: Dr. Lotfi Zadeh.

TEACHING EXPERIENCE

Visiting Professor, Simon Fraser University. Spring, 1989 – Summer, 1989.

Instructor, Summer Linguistics Institute at Stanford University. Summer, 1987. Syntax for Computer Scientists.

Co-Instructor, Stanford University. Spring, 1987. Introductory Semantics. (With John Nerbonne).

Co-Instructor, Stanford University. Spring, 1986. Advanced Syntax. (With Stanley Peters).

Teaching Associate. UCB. Logic for Linguists. Spring, 1981. An introduction to Montague Grammar.

Teaching Assistant, UCB. Typology and Syntax. Winter,1980.

Teaching Assistant, UCB. Introduction to Linguistics. Winter, 1979.

PRESENTATIONS

"Toward a Theory of Locative Reference" (with Lew Creary and John Nerbonne). Unpublished paper delivered at December LSA in San Francisco.