TR-I-0225

# Report on working discussions and ongoing research

Christian BOITET[1]

ATR Interpreting Telephony Research
Sanpeidani, Inuidani, Seika-cho, Soraku-gun
Kyoto 619-02, Japan
xan%atr-la.atr.co.jp@UUNET.UU.NET

1991.9.12

## Abstract

ATR Interpreting Telephony Research has been set up in 1986 to pioneer basic research in the completely new field of Machine Interpretation of spoken dialogues. A small prototype for Japanese-English spoken translation has been demonstrated. Building a prototype for spoken (bilingual) dialogue interpretation still seems to be a long-term goal. Apart from prototyping, ATR-IT is conducting fundamental research in related areas, such as speech processing, AI, algorithms for unification, parallelism, new paradigms for Machine Translation, and discourse/dialogue understanding.

As often in such situations, researchers are encouraged to do basic research, and then asked to produce realistic, nearly preoperational prototypes, two mutually incompatible activities. Other problems, specific to ATR, are that researchers usually stay for less than 3 years, which severely limits the "memory" of the team, and that the proportion of trained linguists is very small, which leads to research being done more on computational problems than on problems specific to MI. Given these circumstances, and the expectations of the scientific and industrial communities, what course could ATR-IT take for the second half of the allotted time of 15 years?

This report has been prepared during a short stay at ATR, on suggestions by MM. Kurematsu, Morimoto and Iida. It contains an account of detailed discussions held with a number of researchers at ATR Interpreting Telephony Research, followed by an analysis of recent research at ATR, based on these discussions and the reading or rereading of relevant literature. Finally, some directions for future basic research directly related to the overall (engineering) goal of MI are suggested.

Keywords & abreviations

Machine Interpretation (MI), Machine Translation (MT), Example-Based MT (EBMT), Dialogue-Based MT (DBMT), Natural Language Processing (NLP), Artificial Intelligence (AI), Bilingual Knowledge Banks (BKBs), Speech Recognition (SR), Speech Synthesis (SR).

---

[1]Visiting researcher from GETA, IMAG, University Joseph Fourier of Grenoble & CNRS, BP 53X, 38041 Grenoble Cedex, France. (boitet%imag.fr@RELAY.CS.NET, xan%frgeta11.bitnet@CUNYVM.CUNY.EDU)

# I. Context of the discussions

The initial plan for the discussions, as been proposed by Mr. Morimoto, is followed in this section. However, the discussions occurred in a slightly different order, because some researchers had urgent work to complete.

## 1. Data processing department

### 1.1. Overview of SL-Trans & interface between Speech Recognition and Analysis.

A demonstration of SL-Trans on 19/8 was followed by several discussions on other topics, some reading, and a discussion with MM. Morimoto and Takezawa on 28/8.

The prototype translates spoken Japanese sentences into spoken English. It is the concatenation of an SR part (HMM-LR followed by "dependency" filter), an MT part (of transfer type, relying on unification-based grammars and dictionaries), and a speech synthesizer (DEC-Talk). The input must be spoken as separate "bunsetsus". Although no context is yet carried over from one sentence to the next, a subsystem has been added after analysis to solve anaphoras.

Although the demonstration as such was quite successful, and its setting up must have represented an impressive amount of work, there are several problems with that prototype in the context of research in Interpreting Telephony, of which the researchers are well aware, and which we discussed at length.

*First, the prototype gives the impression that ATR is researching classical, unidirectional MT, with added speech input/ouput, rather than MI of (bilingual) dialogues.* Given the initial emphasis on AI paradigms, one would have expected some simulation of a human interpreter, who sometimes asks questions to the participants, and makes use of some (maybe shallow) representation of the ongoing total dialogue.

*Second, the coverage of the linguistic data is not realistic* (dictionary of 420 words, with no analysis of compounds, in particular numbers, so important for adresses and phone numbers, and no provision for unknown words, such as proper nouns). Feasibility of Interpreting Telephony (IT), at least for the applications suggested (conference registration, hotel reservation, etc.) should be demonstrated with about 3000 words. Experiments made on the SR part with up to 8000 words have shown that the top 5 candidate bunsetsus produced would contain the correct one in more than 90% of the cases. Hence, the limitation comes from the MT part, and seems to be due to the choice of a strategy almost entirely based on unification.

*Third, there is no deep integration of speech processing and language processing,* as the interface between SR and MT is simply the "most plausible" string of characters delivered by SR. Given the fact that many Japanese companies have developed large-scale MT systems, they all could add an SR and an SS component to their systems and immediately obtain comparable systems of larger coverage, limited only by the SR part.

Perhaps it would be a good idea for ATR-IT to try to construct its next prototype according to the *specificity* of its goal, namely Interpreting Telephony. Comparison with existing large-scale systems for written texts should be avoided whenever possible.

For that, *the prototype would first have to correspond to some realistic situation* (see 3.1 below). It should of course contain the corresponding ergonomic ingredients (e.g., a human bilingual expert might help at the beginning of the dialogue, and be callable if needed; the automatic interpreter should be allowed to conduct metadialogues with each participant, etc.). That would also put ATR in really new territory.

*Second, its internal working should be somehow specific* to the new kinds of problems encountered. As I argued in 1988 [ATR-TR-I-0035], there is not much to be gained from trying to improve the MT part as if classical MT of written documents were concerned. Much more important would be, for example, to connect SR and MT in a tighter way. In the current arrangement, almost all work done by the HMM-LR+filter part is lost, and MT starts almost from scratch. But, in the context of IT, MT should start with a structure like a weighted lattice, with nodes possibly containing the surface structures of the recognized bunsetsus (that is, the structures corresponding to the CFG grammar associated with the LR component).

*Third, overall speed should be a major concern.* For example, it is an error to base analysis on unification, JPSG and the like. While such formalisms may be excellent to formalize and experiment with elegant linguistic descriptions, they are known to be intrinsically far costlier than instantiation-based formalisms (such as attributed grammars used in METAL or in J-SHALT, or ATNs, or even transformation-based analyzers used in HICAT, MU or Ariane). Unification should be used at its proper place, which is, in my opinion, at the most abstract levels of the overall system. Every effort should be made to increase speed (example-retrieval, parallelism of different varieties...).

Despite the above-mentioned problems, *several positive sides of the prototype appear when studying its parts*. The first is the organization of the SR part, which nicely integrates speech recognition techniques (Hidden Markov Models) and some linguistic knowledge (CF grammar for bunsetsus) in the same engine (HMM-LR). Results seem to be quite good, even with 8000 words, and are reported to fare quite well in comparison with other SR systems for Japanese. Other points are mentioned in the appropriate subsections below.

### 1.2. Grammar formalism of Japanese analysis

The grammar actually used in SL-Trans has been presented by Mrs Tomokiyo on 26/8. It has been first developed by MM. Yoshimoto & Kogure, with addition from some work done by Mrs Kume, then reshaped by Mr Nagata, then handed over to Mrs Tomokiyo. It comprises about 20 rules (binary CFG rules, plus f-structure equations, plus some metatules). The dictionary items are also treated as grammar rules. An interesting point is the representation of sentences as pairs <illocutionary force, propositional content>. Note that the grammar is fairly general, and does not yet account for ("ill-formed", but actually used) constructions occurring in real dialogues.

All predicative items (having "valencies") are subject to "subcat slash scrambling". That expression means that, if a lexical item may require up to n arguments, all possibilities are generated statically (16 if n=3: nothing, 1, 2, 3, 12, 13, 21, 23, 31, 32, 123,..., 321). Hence, although the content of each lexical item seems to be quite simple (as compared to that of the MU system, for example), its description is quite long (due to the JPSG formalism), and is often considerably expanded by "scrambling".

It was not possible to discuss with Mr Nagata, who was at CMU at the time. But he seems to follow a promising course by trying to delay unification as far as possible and to eliminate most metarules (see reference). Also, it seems that JIRCO is developing some more extensive dictionary, together with adequate support tools. Then, the JPSG grammatical rules and the lexical database could be used as sources for another formalism, better suited to fast parsing algorithms.

### 1.3. Transfer

This part has been discussed with Mr Suzuki on 26/8. For efficiency reasons, the engine used is a rewriting system for (untyped) feature structures written by Mr Hasegawa, and not RETIF, another rws for (typed) feature structures, on which much work has been done at ATR.

As it is, that engine offers elementary but adequate pattern-matching, and control by global variables and applicability conditions of rules referring to them (environments). The execution of a transformational system uses essentially a top-down recursive traversal of the input f-structure. Intermediate results may be computed from rules by calling a subsystem (defined by some appropriate environment) on a locally constructed f-structure, and using parts of this intermediate result to construct the main result of the rule.

For the moment, the system produces all possible results, with no way to express any preference or to compute any score. That is helpful for grammars designers, but not for a working system. Accordingly, work is under way to extend the system with preferences.

### 1.4. Generation

This part has been discussed with Mr Kikui on 29/8. Here again, previous work (by MM. Ueda and Ogura) was not reused in the prototype, for efficiency reasons. The rewriting system used in transfer has been augmented in order to construct trees from f-structures. It is planned to introduce another rewriting system, for trees annotated with f-structures.

Discussion has centered on the engine and on the linguistic technique for generation. As far as the engine is concerned, it seems that something like GRADE (or ROBRA), adapted to f-structure annotations, could be used as a unique engine for both transfer and generation, instead of several engines. On the linguistic side, transfer and generation could be simplified and clarified by using lexical units in Vauquois' sense (derivational families of lemmas, giving access to some of Mel'tchuk's lexico-semantic functions), rather than mere lemmas.

#### Other remarks

The quality of speech generated is that of DEC-Talk, reading a string of characters. As DEC-Talk offers many ways to control acoustic parameters, it might be a good idea to try a tighter integration with the linguistic generation part in the future. A classical technique to do that is to compute some prosodic and acoustic features during syntactic generation, using the constituent structure as well as other information encoded in decorations (such as illocutionary force, or theme/rheme distinction), and to produce corresponding commands to the synthesizer in the output string.

A very positive aspect, not covered in the preceding subsections, is the development of a speech and dialogue data base, which seems quite huge (700000 words) and well organized.

## 2. Language processing department

### 2.1. Overview of advanced NLP

This part has been discussed with Mr Iida on 27/8. He presented the ongoing research, thereby emphasizing ATR's efforts on dialogue analysis, Example-Based MT, Transfer-Driven MT (see below), and distributed processing. He also underlined the difficulty to get good researchers, enough engineers, and to keep them for sufficient time. That led to a preliminary presentation of the main themes of research envisaged for the next 7-year phase.

The impression from that last part of the discussion was that ATR would like to concentrate on new paradigms for classical, unidirectional MT (EBMT), to extend the current architecture (implemented in the prototpe) to other languages, and, perhaps as a main point, to research the use of parallelism for integrating various sources of knowledge (text, speech, vision,...). Mr Iida said that this was due to external circumstances, which would be explained later. However, this impression was somewhat corrected in a later meeting (see section 3 below).

## 2.2. TDMT & EBMT

This part has been discussed with Mr Furuse on 30/8. The term "transfer-driven" is somewhat unfelicitous, given the number of "transfer-based" systems. The idea here is rather "Least-Effort-Based MT". That means that translation should be attempted at the lowest possible level, if recognition (in lower levels) or analysis (in higher levels) reaches a certain score. Three levels are envisaged: fixed sentences, string patterns, tree patterns. The first two levels have been implemented for a demonstration system, which makes use of ATR's data base.

An interesting point is the computation of distances between word pairs, based on a 3-level thesaurus of Japanese, computerized by ATR for this purpose. Also, it is interesting to note that a high percentage (≈30%) of utterances in ATR's dialogues seem to be "canned", as well as their translations (e.g., "moshi moshi" -> "hello"). The demonstration itself was quite impressive.

That work is influenced by research done at Kyoto University and at BSO (see references), with which ATR entertains scientific links. At the third level, it must make use of a data base (BKB, in BSO terminology) of previously analyzed "bitexts", with correspondences between subtrees. It is not yet clear at all whether that is possible for pairs of distant languages such as Japanese and English, at least if the structures are "concrete" (as opposed to "abstract") dependency trees. Also, future research should be done on the overall engineering cost of such techniques: in addition to some classical, large-scale analyzer and generator, it seems necessary to construct, update, store and access huge BKBs.

## 2.3. Anaphora resolution

This part has been discussed with Mr Dohsaka on 2/9. The research focusses on the resolution of zero pronouns referring to persons in Japanese dialogues. That is quite important for correct generation of full nominal expressions (e.g., "musuko", "musuko san"...). Normally, this resolution makes use of constraints produced by the interpretation of previous utterances in the dialogue. A version limited to the context of a sentence has been integrated into the current prototype.

Anaphora resolution is obtained through abductive inference, realized using a kind of constraint programming. It would perhaps be useful for ATR-IT to develop or acquire a general engine for constraint programming, in order to integrate work on other types of constraints.

## 2.4. Discourse analysis & its application to speech processing

This part has been discussed with Mr Yamaoka on 3/9. The idea is to define plans, to recognize them as the dialogue is progressing, and to use that recognition to predict the possible (kinds of) next sentences. The demonstration system runs on a dozen of dialogues taken from ATR's data base (in Japanese, but there is also an English mode of presentation).

It may be suggested to consider the definition of plans and sub-plans in a grammatical framework, and to use (adaptations of) known algorithms to speed up the recognition/prediction process. In turn, the technique used in the current program is somewhat more general than what is done in usual syntactic analysis: a new utterance may "fill" a waiting preterminal which is not "at the top of the stack" (there may be several stacks, as there may be several concurrent – and current – partial interpretations).

## 2.5. Parallel parsing

This part has been discussed with Mr Neuhaus on 23/8, and with Mr Myers on 2/9. Three kinds of parallelism are envisaged: distributed processing (e.g., with several Sparcs on a network), parallel programming (Sequent), and massively parallel programming (Connection Machine).

Parallel parsing is now being attempted on the Sequent, which is equipped with a dozen of processors sharing a large core memory, and runs Allegro CLiP, a parallel version of Common Lisp. It is not clear whether significant improvements can be obtained on that architecture, with the current unification-base grammatical framework, as, even using a dynamic programming algorithm such as CKY for the context-free (CF) base, the number of f-structures grows exponentially.

Hence, if unification is performed at each rule application, no speed-up can be maintained (for the same upper bound on sentence length) if the grammar grows (remember that the grammar also contains the dictionary in the current framework). But, if unification is delayed until full analysis trees relative to the CF base are obtained, there can still be an exponential number of them. What seems to be needed is to introduce in the rules conditions and actions on *bounded* attributes. These conditions and actions would normally be simplifications of the f-structure equations.

Distributed processing has been suggested for the next prototype. This will call for new research on an appropriate shared data structure, to be enriched by each module, and perhaps only partially represented in each of them. Or, if the central scheduler supports the one and only copy of that structure, would that kind of implementation be very different in principle of one using the Sequent?

Massive parallelism is researched in cooperation with the CMT (CMU). This direction is extremely interesting, in that it calls for really new architectures of an MI system [Kitano91, Tomabechi91].

### Other remarks

It was somewhat difficult to understand clearly the difference between the DB and the NLP departments. As the third department is concerned with research and prototyping in one domain (speech processing), I first assumed that the NLP department was supposed to do basic research and prototyping on aspects specific to NLP, and possibly AI, while the data base department would do the same on all kinds of data bases, including dialogues, texts, grammars, dictionaries, and computer tools (specialized languages, environments) reusable by all researchers.

As a matter of fact, these denominations seem to have staid only for historical reasons, and it seems that the data base department is doing more actual language processing than the language processing department. I then supposed that the first department was responsible for prototyping a complete system, and do applied research, while the second had to take care of fundamental, basic research. But that again does not seem quite true. Perhaps it would be useful to find a clear definition, which would in turn help in defining, and then assessing, the various research themes.

## 3. Plan for next project

This topic has been discussed with Dr Kurematsu on 20/8, with Mr Iida on 27/8, and with MM. Morimoto and Iida on 31/8. It seems that the different organisms and firms supporting ATR have somewhat conflicting goals. Some would like ATR to do only basic research, some are asking for "working" prototypes, some would like to avoid any concurrence with their proper research labs (NTT) or with their proper MT technology (computer makers), etc. In addition, the next project should (at least appear to) be new, while continuing in the same direction (Interpreting Telephony).

Although the overall picture is not very clear, it seems that the current plan for the next project aims essentially at more basic research, with objectives such as experimenting new paradigms for MT, and treating "natural" utterances. A related goal is to attract good researchers and engineers for longer periods. However, further development of a prototype (for MI) seems to be envisaged, as well as further extension of the current prototype to other language pairs (such as Japanese-German).

# II. Analysis of some aspects of recent research

That analysis is based on the discussions reported above, and on the documents listed in the reference list. I have tried to read as many as possible of the documents published in English by ATR-IT since 1988, as well as other pertinent available sources. However, I may have completely missed some lines of research, in particular if all references are in Japanese, but also if there are conference papers in English, which I did not read before, and which have not been published as ATR reports.

The research themes are somewhat arbitrarily divided in seven classes, presented in alphabetical order.

## 1. Artificial intelligence (architectures, methods, systems)

Of course, there are AI aspects in most of the work mentioned in the following sections. Here, we consider only work aiming at AI "proper", that is, having general applicability in AI, and not only in MI, or even NLP in general. For instance, we don't refer here to research on plan recognition and abductive reasoning [Yamaoka&Iida91, Dohsaka91], conducted specifically in the context of dialogue understanding.

Given that restrictive definition, research done on AI is quite substantial, but does not appear to be a priority of ATR-IT. Research work on AI-oriented architectures is presented in [Myers90] and [Tomabechi91]. As far as tools are concerned, there seems to be only one general purpose AI system, namely the ATMS (truth maintenance system) documented in [Myers89].

## 2. Linguistics & (man, machine) translation

[Stanwood&Suzuki90] is the only ATR reference in the following bibliography. As a matter of fact, there have been papers at conferences, for example at EACL-89 and COLING-90, mostly concerning the computation of speech act types and honorific expressions, by Mr Yoshimoto, Mrs Kume, and others. More linguistic studies seem to be necessary, but, as mentioned in the introduction, ATR can only have a very small proportion of linguists.

The preparation of the data base cannot in itself be seen as linguistic research, but it offers a rich base from which basic research (e.g., dialogue grammars, regularities behind "ill-formed" spoken utterances, contrastive studies), based on actual data, could be performed efficiently.

More applied research would also be possible. For example, there are numerous computerized lexical resources in Japan, some of them no doubt accessible to ATR, like those of EDR, NTT, KDD, JICST... Work in computational lexicography could lead to methods and tools for converting information contained in existing lexical data bases or computerized bilingual dictionaries into a format usable at ATR.

## 3. Machine Interpretation

Several domains are researched here, namely integration of speech processing and language processing [Kurematsu&al91, Morimoto&al91], resolution of anaphoras [Dohsaka91], plan recognition and dialogue understanding [Yamaoka&Iida91], identification of specific problems [Myers&Toyoshima90], and exploitation of the data base [Huber91].

Although the explored topics are interesting *per se,* there is a feeling that researchers have not yet really tried to tackle issues *specific* to the MI of dialogues. It has already been mentioned that no work seems to have been done on the analysis of real telephone dialogues using human interpreters, although such data seems to have been collected in the framework of a research contract with SRI. Rumour has it that up to 90% of the time is spent in metadialogues between the interpreter and the callers. If true, that would be extremely important for the general design of an MI system.

Even if one supposes that an MI system could be a "black box", the dialogues would be bilingual. But all work to date has centered on the dialogues in Japanese only. It is quite difficult to imagine situations where one would like to translate such dialogues over telephone lines. It is also not clear whether the problems most important in monolingual dialogues are also the most important in bilingual dialogues.

## 4. Machine Translation

Research is done on the overall architecture of MT systems, on their linguistic base, and on suitable parallel implementations.

Research on Example-Based Machine Translation (EBMT) is now becoming popular, and ATR's data base gives researchers a considerable amount of primary data. There are at least two challenges for the future. First, prepare a large enough BKB (with analyses and correspondences), and find methods to (piecewise) match a result of analysis (f-structure or decorated dependency tree) against the source language part of the BKB. Second, determine the domain of applicability of that technique. For example, is it possible to get a BKB large enough for obtaining good translations in hotel reservation dialogues?

Research on the linguistic base has already shown the importance of structuring the lexicon(s) with lexico-semantic functions *à la* Mel'tchuk. It seems quite important to try and apply that theory to Japanese, at the same time trying to find just how many such functions should and could realistically be integrated in a lexical data base for MT and MI. That domain seems relatively new in Japan.

Research on parallel implementations has begun because of the slowness of unification-based parsers and the need for very high speed in interpretation. In turn, the goal of parsing, understanding and generating in (quasi-)real-time will no doubt give rise to new linguistic architectures for MI.

[Morimoto&Iida91] rightly insists on using parallelism to integrate all aspects specific to MI (dialogue, spontaneous speech...). As a matter of fact, it would not be so interesting to study parallelism in the context of classical, text-oriented MT systems, because fast enough solutions already exist. Relatively low quality MT systems adequate for information gathering are already available on portable workstations (Sharp, Toshiba), relatively high quality systems for professional post-editors run fast enough on minis and on workstations, and, for very high quality in restricted sublanguages, simple LBMT systems can be operated on low cost PCs under severe time constraints (METEO runs on a micro since almost 10 years).

## 5. Neural nets

No discussion has taken place on that topic. However, that domain seems to be one of the strongest in ATR. Researchers on that theme seem to function as a team, going together in the same direction.

Progress is impressive. Three years ago, excellent results were obtained for recognition of very small phoneme sets (e.g., BDG), but it was not at all clear how to upscale the technique and produce a large NN recognizing all phonemes of Japanese. That has been achieved, and a lot more. It seems that TDNNs, coupled with LR mechanisms, can now be usable for vocabulary-independent recognition of phonemes [Sawai91].

## 6. Specialized languages & environments for linguistic programming

Relatively few research on that topic is conducted at ATR, because most researchers are computer scientists who like to develop their own prototypes from cave to roof. However, that makes it difficult to experiment with variations in a complete prototype, and the minority of linguists cannot be as productive as it could.

For example, the analyzer has no debugging environment and no informative diagnostics facility, so that, in case of failure, the linguist has to try again, on pieces of the input, in order to guess the probable source of the problem.

Note that it would not be so costly to take the specialized tools created by some researchers (like rewriting systems, parsers...), to make them independent of the underlying implementation language (so that more efficient implementations could be prepared transparently), and to hide the underlying file system by creating a logical environment. The construction of tools for lingware development may not be "basic" research, but is basic for making real advances. Nevertheless, it is a recognized research area, with publications in journals and conferences (see for examples communications on D-PATR, Kimmo, UP, GRADE, METAL, Ariane...). Another problem is to ensure some continuity in the support (maintenance, updating, documentation).

### 7. Unification

A lot of interesting work has been done at ATR on that particular point. Some is concerned with finding faster algorithms to unify classical f-structures, and some with the extension to typed feature structures, with or without multiple inheritance.

That line of research has produced some interesting theoretical results (operations on typed f-structures, notably) as well as complete implemented environments. It is an example of fruitful continuity (Kogure, Zajac, Nicolas, Emele, Nagata, Tomabechi, Neuhaus, Furuse, Iida...).

### Summary

With due regard to my possible misinterpretations and overlooks, the remarks above may be summarized along four main lines:

- Research themes don't appear to be really centered on MI of telephonic (human) dialogues. They sometimes duplicate research pursued much more intensively at other places. On the other hand, some problems specific to the MI of dialogues are not covered by current research.
- Teamwork is quite rare: too often, research themes seems to appear with individual researchers, and to disappear when they leave.
- To an external observer, the research themes for which ATR-IT is best known are unification, neural nets, and parallelism.
- A very positive point is that all researchers try to implement their ideas, and sometimes come up with impressive demos. The quality and quantity of available harware is remarkable.

## III. Possible basic research themes linked with application-oriented prototyping

ATR-IT has chosen to pioneer a new domain. In my opinion, basic research should consist in really exploring that domain, and especially its new aspects, rather than enter in concurrence with other labs or firms on domains which have been largely explored, or tackling problems not directly concerning MI.

For that, prototyping in the context of realistic situations should be a major objective in the future. *That objective should be pursued as an individualized project.* Until now, it has more or less been seen as a secondary goal, achievable by simply putting together the available results of basic research. In NLP in general, and MI (or MT) in particular, that does unfortunately not seem to be feasible. The Eurotra project is a case in point.

Basic research themes could be more centered toward ATR's central objective, especially if the prototyping activity helps in selecting the most important problems, and in testing tentative solutions by incorporating them in (copies of) the prototype. Let us illustrate this idea by mentioning some interesting and specific basic research themes directly related to Interpreting Telephony.

# 1. Interpretation of spoken dialogues (2 source/target languages)

## 1.1. Situations

We are looking for situations where two humans *need* to communicate over the phone, using mainly speech and not a keyboard with a video screen. In this light, conference registration is not very realistic. It is extremely rare to phone to a conference office, as almost everything is done by mail or e-mail, and the task is adequate for a videotext kind of automation.

Hotel reservation and car rental have also been suggested. Let us discuss them briefly, first in the context of international communication. If a Japanese goes abroad, he most probably knows enough English to book a hotel room, and the hotel personnel are also likely to speak enough English. If he does not know English, he probably books through a travel agency, which is in contact with a small number of large hotel chains having English speaking personnel. In most cases, because of time difference (and perhaps cost), reservation will be handled by telex or fax. About the same can be said of a foreigner travelling to Japan, and of the car rental situation in both cases.

The situation is different if the foreigner is residing in the other country, or travelling there for a long time. Such people are likely to prefer inexpensive small hotels and local car rental agencies, where nobody is bilingual. Dr Kurematsu has suggested the term of *"assistance to reservations by foreign travellers"*. Other types of reservations can be envisaged (shows, sports events, restaurants...).

Another kind of situation is *"arrangement of an appointment or a pickup"*. In many cases, the traveller is not in his office or home, but has to use some public phone. For example, his flight is delayed, and he phones from some airport to make new arrangements, or he arrives at some train or bus station in town, and has to tell his correspondent where to pick him up, etc. International as well as national communications are concerned.

It remains to be seen whether these situations make sense economically. Is there already some provable demand? KDD already offers human interpretation: what kind of conversations are translated? Is there some comparable service offered by NTT or other telephone companies, in Japan and abroad?

Suppose that the situations described above are realistic goals. The MI system can be designed either as a simulated interpreter, who "brokers" the conversation, thereby interacting with the participants, or as a "black box".

## 1.2. Machine Interpreter as broker

If the MI system acts as a human interpreter, several specific topics could be researched.

*First, not only dialogues, but also metadialogues must be modelled.* Very often, the participants will speak *to* the MI system, thereby employing direct style to speak to the MI system ("please repeat", "did he say 30th or 13th?"), and indirect style to indicate they are speaking to their correspondent ("please tell him that..."). In the same vein, it is likely that the MI has to change familiar expressions into more respectful ones.

*Simulating an intelligent interpreter* is in itself a very interesting and complex endeavour, certainly leading to AI-oriented research topics. It seems that the MI system should be able to create and maintain models of the correspondents, of their dialogues, of the two metadialogues, and of the progression of the task at hand.

It is also interesting to try to *determine the "best level" of understanding of the system.* After all, it is assumed that two intelligent humans are talking, and they may be far better than any system when it comes to ask the right disambiguating question in the current context!

On the MT side, *an open question is whether EBMT can be integrated in the framework of an understanding system.* What can be understood if a canned translation has been retrieved, and no complete analysis has been made? In order to understand, it seems that a system needs to relate some rich internal representation of the utterances with an abstract representation of the world of reference, the actors (including, but not limited to the correspondents), and the ongoing dialogue and its associated pragmatics. Perhaps the system could analyze further after having translated, and integrate the results, while continuing to listen to the next utterance?

On the speech side, it seems interesting to try to *produce a voice which appears to be the same in the two languages, and which is clearly distinguishable from the voices of the two participants.* (A research on Japanese-English voice conversion has already been carried out by M. Masanobu Abe). It would also be interesting to investigate whether some phonetic parameters can be modified to make the difference between dialogue and metadialogue more sensible.

### 1.3. Machine Interpreter as black box

That architecture may be desired, in order the reduce the length of the communications. Without meatadialogues enabling the system to clarify some points, and sometime even to "negotiate" the input, no very high quality can be expected. Also, very complex and time consuming deep understanding processes don't seem to be desirable.

On the other hand, that is the type of architecture where *example-based translation* would seem most appropriate. Can it be used effectively for this kind of translations? If yes, can it be adapted to the speech level? In particular, can good prosody be generated if there is no direct match? Are the results really superior to those of systems based on dictionaries and grammars?

Then, there are *questions of control and ergonomy.* For instance, should the system send to the Japanese speaker a (Japanese) version of what is generated in English, with zero-pronouns replaced by the same explicit references as in the translation?

On the speech side, in contrast to the preceding architecture, one could try to *produce voices as similar as possible to those of the two correspondents,* in addition to a neutral voice for the system.

## 2. Interpretation of spoken input (1 source language)

### 2.1. Situations

If the input text preexists in written form, it does not seem to make sense to read it aloud to a machine in order to have it translated. If it is printed, there are quite good OCR systems. In the case of a handwritten text, keying in is not so long, but careful spoken input (discontinuous speech) might be an alternative. However, it is difficult to imagine a realistic application.

A possible class of situations is issuing *multilingual warnings.* Take earthquakes, for example. The person in charge of issuing warnings would first utter a warning (in Japanese), to be broadcast immediately, then create another one, for the benefit of foreigners, to be translated from Japanese into English, and possibly other languages. Interaction with the MI system seems possible, and written input is not likely. We have a situation of Dialogue-Based MT (DBMT) with spoken input, and no possibility of multilingual generation from an expert system.

In the same vein, there are *announcements at international events* (such as Olympic Games, international fairs, exhibitions, conferences...). They may be broadcast on electronic panels, through loudspeakers, or over the radio, and they can be made available on videotext, or through telephone. In the future, such announcements might be generated by some kind of expert system, in the spirit of the work recently done in Canada by Kittredge & Polguère on multilingual generation of weather reports. But that does not seem to be an immediate concern, and, in the meantime, human announcers will continue to do the job from some cubicle. Note that both situations may coexist: Canadian weather bulletins are still produced locally by humans, sent over telex lines, and automatically translated by METEO, although some general situation reports are now generated automatically.

Another application could concern the *translation of spoken information.* Often, medias get tapes with spoken interviews, speeches... in foreign languages, say English, which they want to translate (for subtitles or overdubbing). To treat such difficult input (large coverage, bad acoustics, continous speech, several speakers, and ill-formed utterances) is clearly beyond the possibilities of MI, at least in the foreseeable future.

But a (monolingual) native speaker of English could listen to the tape, and produce a clear rendering, possibly in discontinous speech, with good acoustics, well-formed utterances, vocal marks (to distinguish original speakers, to delineate sentences), etc. A spoken MT system could then produce a "good enough" raw translation.

### 2.2. With a source text

If there is an initial linguistic formulation, we call it, by analogy with usual translation, the source text. That can be the case in the three kinds of situations described above. A distinctive possibility is to "negociate" that input text with its author, so that it falls within the analysis capabilities of the system, and to "clarify" it, in order to interactively solve all ambiguities not solved by the language processor.

Considerable research will be necessary to find ways to *generate that kind of man-machine negotiation and clarification dialogue in a spoken and ergonomically acceptable way.*

### 2.3. Without a source text

"Translation without a source text" is a term coined at UMIST by Pr Tsujii and his colleagues. The idea is that the system helps the human in building a message, which is then translated, and optionnally generated in the source language for control and bookkeeping purposes. That can be the case for warnings and for announcements. This "paradigm" could be used even if there were no "ontology", or a very shallow one. In other words, it can be based almost completely on linguistic knowledge.

The main problem here seems again to find ways to interact with a user through speech in a manner as user-friendly as what can be done using a screen, a mouse and nice graphics. But the nature of the dialogue is very different from the preceding situation.

## 3. Other (common) themes

In all imaginable situations for MI, the input is considerably more noisy than in MT (of texts). Basic studies on ways to really handle that situation at all levels of processing offer interesting perspectives. *"Ambiguous programming"* techniques should be developed. The idea is to represent all levels of interpretation, and all ambiguities, in a factorized way, but to write the biggest part of the linguistic "programs" *as if* there were no ambiguities, and to handle ambiguities in separate parts, as problems identifiable by patterns in the underlying data structure (e.g., a weighted layered lattic).

That is reminiscent of "non-deterministic programming", in Prolog for example, where one programs the solutions independently, *as if* there were no non-determinism, and controls that non-determinism separately (e.g., using a "cut").

As mentioned before, there are many possible themes for *research in linguistics proper,* most notably in structuring the lexicons, defining adequate abstract and concrete representations, studying correspondences between them in bitexts, etc. Also, aspects concerning lingware engineering are quite essential.

*Multimodal interaction and other ergonomic aspects* are also important themes, in that acceptability of an MI system might well depend more on them than on the linguistic quality of the translations.

For speech generation, an interesting question is *how to compute prosody, energy, pauses, rythm, etc., from detailed structured linguistic representations*. Phoneticians and grammar writers should cooperate to build phonetic oriented generators.

In the context of the media application, the MI system could produce a medium-quality first draft, in written or spoken form. Here, delay is the main constraint, not translational perfection (if that exists). Interesting research could be conducted on the feasibility of *spoken editors* for oral postedition.

How to use *parallel architectures* is also an important theme. As said before, that line of research may well contribute to the discovery of new "paradigms" for expressing various types of knowledge, and not only for computing structures derived from classical representations of knowledge.

## Conclusion

Research conducted at ATR-IT has tackled various themes, inspired by the long-term goal of Machine Interpretation of telephone (human) dialogues. That research could however be more centered on MI, with more intensive research being done on fewer, but more specific themes.

For the future, it might be a good idea to try to build a full-fledged prototype for some realistic application evidently requiring speech input. That prototyping activity should be autonomous, that is, it should not depend directly on results from basic research.

In turn, basic research could benefit from the prototyping activity as a source of inspiration for real and new problems, and as a tool for experimenting with new ideas in a realistic setting, and not, as is too often the case in NLP, on fabricated examples.

Such cross-fertilization would be facilitated if each researcher contributed for a certain fraction of his time to the prototype project, while spending the rest on his personal basic research theme.

## Acknowledgements

I would like to express my gratitude to Dr Akira Kurematsu, president of ATR Interpreting Telephony Research Laboratories, for his kind invitation to come again to ATR. Further thanks are due to him as well as to MM. Tsuyoshi Morimoto and Hitoshi Iida, for having spent much time helping me to put this report in perspective with ATR's overall goals and constraints. Last, but not least, hearty thanks should go to Osamu Furuse, Gen'ichiro Kikui, Kohji Dohsaka, Peter Neuhaus, John Myers, Masami Suzuki, Toshiyuki Takezawa, Hideto Tomabechi, Mutsuko Tomokiyo, Noriyushi Uratani, and Takayuki Yamaoka, for interesting discussions and presentations.

-o-o-o-o-o-o-o-o-o-o-

## References

The following list is divided into the (somewhat arbitrary) seven categories used in section 2 above. In each category, references are arranged by dates (most recent first) and by authors. A separate document presents the same list, with more or less detailed comments on each reference.

### 1. Artificial Intelligence (architectures, methods, systems)

[1] KITANO Hiroaki (1991)
Massively Parallel Artificial Intelligence and Its Application to Natural Language Processing.
FGNLP'91 International workshop, ATR, Kyoto, April 1991, p. 87–106.

[2] TOMABECHI Hideto (1991)
MONA-LISA: Multimodal Ontological Neural Architecture for Linguistic Interactions and Scalable Applications.
FGNLP'91 International workshop, ATR, Kyoto, April 1991, p. 49–67.

[3] MYERS John K. (1990)
A Project Report on NP: an Assumption-Based NL Plan Inference System that uses Feature Structures.
Proc. COLING-90, Stockholm, Aug. 1990, 428–430.

[4] MYERS John K. (1989)
An assumption-Based Plan Inference System for Conversation Understanding.
WGNL Meeting of the IPSJ, Okinawa, NLC89-24, 30/6/1989, p. 73–80.

[5] MYERS John K. (1989)
The ATMS Manual. Version 1.1.
ATR report TR-I-0074, Feb. 1989, 85 p.

## 2. Linguistics & (man, machine) Translation

[6] STANWOOD Ryo, SUZUKI Masami (1990)
Some Computational Applications of Lexical Functions.
ATR report TR-I-0179, Aug. 1990, 52 p.

[7] CHUQUET Hélène, PAILLARD Michel (1987)
Approches linguistiques des problèmes de traduction anglais <-> français.
Éditions Ophrys, Paris, 451 p.

[8] MEL'TCHUK Igor A., PERTSOV Nikolaj V. (1987)
Surface Syntax of English. A formal model within the Meaning-Text framework.
J. Benjamins, Amsterdam, 1987, 526 p.

[9] MEL'TCHUK Igor (1984)
Dictionnaire explicatif et combinatoire du français contemporain.
Presses de l'Université de Montréal, 1984, 172 p.

## 3. Machine Interpretation

[10] HUBER Dieter (1991)
A Bilingual Database for Automatic Spoken Language Interpretation between Japanese and English.
ATR report TR-I-0196, Feb. 1991, 70 p.

[11] KUREMATSU Akira, IIDA Hitoshi, MORIMOTO Tsuyoshi (1991)
Language processing in connection with speech translation at ATR Interpreting Telephony Research Laboratories.
Speech Communication 10 (1991), 1-9, North Holland.

[12] MORIMOTO Tsuyoshi, SHIKANO Kiyohiro, KOGURE Kiyoshi, IIDA Hitoshi, KUREMATSU Akira (1991)
Integration of Speech Recognition and Language Processing in a Japanese to Englih Spoken Language Translation System.
IEICE Transactions, vol. E 74, N° 7, July 1991, 1889-1896.

[13] YAMAOKA Takayuki, IIDA Hitoshi (1991)
Dialogue Interpretation Model and its application to next utterance prediction for spoken language processing.
Proc. EUROSPEECH-91, Genoa, 1991, 4 p.

[14] DOHSAKA Kohji (1990)
Identifying the Referents of Zero-Pronouns in Japanese based on Pragmatic Constraint Interpretation.
Proc. ECAI-90, Stockholm, 1990, 6 p.

[15] MYERS John K. (1990)
A Design for a Disambiguation-Based Dialog Understanding System.
ATR report TR-I-0189, Nov. 1990, 116 p.

[16] MYERS John K. (1990)
Methods for Handling Spoken Interruptions for an Interpreting Telephone.
WGNL Meeting of the IPSJ, NLC90-3, 21/5/90, p. 17–24.

[17] MYERS John K., TOYOSHIMA Takashi (1990)
Known Current Problems in Automatic Interpretation: Challenges for Language Understanding.
ATR report TR-I-0128, Jan. 1990, 24 p.

[18] BOITET Christian (1988)
Representation and computation of units of translation for Machine Interpretation of spoken texts.
ATR report TR-I-0035, Aug. 1988, 41 p.
(& Czech Journal of AI, 1989).

## 4. Machine Translation

[19] BOITET Christian (1991)
Twelve Problems for Machine Translation.
Proc. Int. Conf. on Current Issues in Computational Linguistics, USM, Penang, June 1991, 12 p.

[20] FURUSE Osamu, SUMITA Eiichiro, IIDA Hitoshi (1991)
Building Transfer Knowledge from a Bilingual Spoken-dialogue Corpus: Realization of Transfer-Driven Machine Translation.
ATR, internal paper, 1991, 31 p.

[21] SADLER Victor (1991)
The Textual Knowledge Bank: Design, Constructions, Applications.
FGNLP'91 International workshop, ATR, Kyoto, April 1991, p. 17–32.

[22] SATO Satoshi (1991)
Example-Based Translation Approach.
FGNLP'91, International Workshop, ATR, Kyoto, July 1991, p. 1–16.

[23] SUMITA Eiichiro, IIDA Hitoshi (1991)
Experiments and Prospects of Example-Based Machine Translation.
ATR, internal report (for ACL conference?), 8 p.

[24] WHITKAM Toon (1990)
ABMT for Text and Dialogue: a preliminary assessment of its potentials.
ATR report TR-I-0165, Aug. 1990, 32 p.

[25] NAGAO Makoto (1989)
Machine Translation. How far can it go?
Oxford University Press, New-York, Tokyo, 1989, 150 p.

[26] SADLER Victor (1989)
Working with Analogical Semantics.
Disambiguation techniques in DLT.
Foris, Dordrecht, 1989, 256 p.

[27] PAPEGAAIJ B.C. (1986)
Word Expert Semantics. An Interlingual Knowledge-Based Approach.
Foris, Dordrecht, 1986, 254 p.

## 5. Neural networks

[28] SAWAI Hidefumi (1991)
Connectionist Large-Vocabulary Continuous Speech Recognition.
ATR report TR-I-0209, March 1991, 47 p.

[29] ELMAN Jeffrey L. (1989)
Representation and Structure in Connectionnist Models.
CRL-TR-8903, Center of Research in Language, Univ. of California, San Diego, Aug. 1989, 26 p.

[30] SAWAI Hidefumi, WAIBEL Alex, HAFFNER Patrick, MIYATAKE Masanori, SHIKANO Kiyohiro (1989)
Parallelism, Hierarchy, Scaling in Time-Delay Neural Networks for Spotting Phonemes and CV-Syllables.
ATR report TR-I-0090, July 1989, 21 p.

[31] WAIBEL Alex (1989)
Connectionist Large Vocabulary Word Recognition.
ATR report TR-I-0120, Oct. 1989, 16 p.

[32] HAFFNER Patrick (1988)
DyNet, a Fast Program for Learning in Neural Networks.
ATR report TR-I-0059, Nov. 1988, 28 p.

## 6. Specialized Languages & environments for linguistic programming

[33] HASEGAWA Toshiro (1991)
A Rule Application Control Method in a Lexicon-Driven Transfer Model of a Dialogue Translation System.
Proc. ECAI-90, Stockholm, p. 336–338.

[34] KIKUI Gen'ichiro (1991)
The Sentence Generator of ATR Interpreting Telephone System.
ATR, preliminary manuscript, April 1991.

[35] NAGATA Masaaki (1991)
Efficient Unification-Based Grammar Parsing with medium-grained CFG rules and late unification.
ATR, internal paper, Jan. 1991, 16 p.

## 7. Unification

[36] NEUHAUS Peter, FURUSE Osamu, IIDA Hitoshi (1991)
Unification-Based Parsing on Increasing Levels of Parallelism.
WGNL Meeting of the IPSJ, NLC91-9, 19/9/91, 8 p.

[37] TOMABECHI Hideto (1991)
Quasi-destructive Graph Unification.
ATR report TR-I-0198, Apr. 1991, 14 p.

[38] KOGURE Kiyoshi (1990)
Strategic Lazy Incremental Copy Graph Unification Method.
ATR report TR-I-0129, Jan. 1990, 28 p.

[39] CARTER David (1989)
Efficient Disjunctive Unification in a Bottom-up Shift-Reduce Parser.
ATR report TR-I-0124, Nov. 1989, 27 p.

[40] EMELE Martin, ZAJAC Rémi (1989)
Multiple Inheritance in RETIF.
ATR report TR-I-0114, Sept. 1989, 48 p.

[41] EMELE Martin, ZAJAC Rémi (1989)
RETIF, a rewriting system for typed feature structures.
ATR report TR-I-0114, Mar. 1989, 39 p.

[42] NICOLAS Yves (1988)
Pragmatic Extensions to Unification-based Formalisms.
ATR report TR-I-0054, Nov. 1988, 66 p.

[43] ZAJAC Rémi (1988)
Typed Feature Structures II. The language and its implementation.
ATR report TR-I-0055, Dec. 1988, 40 p.

-o-o-o-o-o-o-o-o-o-o-o-o-