音声言語情報処理　4－3
（1994. 12. 15）

Towards More Robust, Fault-Tolerent and User-Friendly Software Integrating Natural Language Processing Components.

Hervé Blanchon and Kyung-Ho Loken-Kim
ATR – Interpreting Telecommunications Research Labs
2–2 Hikari-dai, Seika-cho, Soraku-gun
Kyoto 619-02, Japan
Blanchon@itl.atr.co.jp
kyungho@itl.atr.co.jp

Abstract

Natural language processing components have been introduced in a number of software. These software range from natural language processing systems (as Machine Translation) to systems where natural language is a communication modality with the user. In this paper we are going to establish the soundness of integrating a clarification module in these systems to make them more robust, fault-tolerant and user-friendly. We will then give an account of a first interactive clarification module for French input which has been developed at the GETA Laboratories in France in the framework of the LIDIA project of Dialogue-Based Machine Translation. Finally, we will describe ongoing work at ATR-Interpreting Telecommunication Research Laboratories on the adaptation of the previous clarification module to English input and the development of experiments in the setting of the Environment for Multi-Modal Interaction.

Key Words

Natural Language Processing, Interactive Clarification, Dialogue-Based Machine Translation, Machine-Aided Interpersonal Communication, Multimodal Systems.

Introduction

Spoken or written natural languages are going to be more and more used as interaction modalities between human users and interactive software. Natural language processing techniques do not allow a really robust, fault-tolerant and user-friendly use of these modalities. The issues related to natural processing techniques can be divided in two categories: the textual input issues and the spoken input issues.

As far as the textual input is concerned, some problems can occure with the the spelling or the syntactic consistency. However, the most important difficulty is the ambiguousness of the input to be analyzed. In fact, ambiguities can arise even if the domain of the utterance is well defined and many knowledge sources are involved in the analysis process.

As far as the spoken input is concerned, the first problem concern speech to text transformation. Then, difficultes related to textual input arise. In speech to text transformation the difficulties stem mostly from two problems: segmentation and variability. Moreover, even when the segmentation has been done without any ambiguity, the problem of homophony has to be solved.

For the use of natural language to be more robust, fault-tolerant and user-friendly we propose to integrate a clarification module as a component of every concerned interactive software. The role of such a module is to help the recognizer (for speech) and the analyzer (for text) to produce an unambiguous representation of the user's input corresponding to the his intention.

In our opinion, clarification has not been yet studied as a core research framework and that's what we want to promote. Indeed many ambitious projects are using natural language as a communication modality between the system and the user. Then, we think that the need for clarification framework will be more and more felt.

As a first step in this work we rely on:

- a first experiment with a clarification module in the framework of the LIDIA project (Large Internationalisation of Documents through Interaction with the Author) at the GETA Lab. (Study Group for Machine Transtalion),
- firsts results of the MIDDIM (Multimodal Interactive Disambiguation) joint research project between ATR-ITL (ATR-Interpreting Telecomunication Research Labs) and the GETA Lab.,
- the EMMI (Environment for Multi-Modal Interaction) simulator of Machine-Aided Inter-Personal Communication at ATR-ITL.

In this paper we will first introduce what we call the clarification framework. We will then give an account of a first interactive clarification module for French input which has been developed at the GETA Laboratories in France in the framework of the LIDIA project of Dialogue-Based Machine Translation. Finally, we will describe ongoing work at ATR-Interpreting Telecommunication Research Laboratories on the adaptation of the previous clarification module to English input and the development of experiments in the EMMI setting.

I. The clarification framework

I.1. The need for clarification

Natural language (spoken or written) is seen as a really attractive modality. Speech is attractive because, as stated in [Kay, *et al.* 1994, pp. 110-111]: speech requires no training, speech is fast, speech requires little attention. Text also can be attractive when the utterances are short, when speech is not mandatory and when the use of speech can be very annoying for the entourage of the user. Recent foreseen applications using natural language interface include multi-modal drawing tools [Caelen 1994 ; Hiyoshi & Shimazu 1994 ; Nishimoto, *et al.* 1994], on line travel information [Goddeau, *et al.* 1994] and more generally on line information retrieval [Haddock 1992], oral control systems and finally Machine-Aided Inter-Personal Communication (MAIPC) [Kay, *et al.* 1994 ; Morimoto, *et al.* 1992].

Nevertheless, natural language input has always been, and will continue to be, handled with great difficulties by computers. At least right words, syntactic structures and surface semantics features used in the input must be recognized. As natural language, either spoken or written, is highly ambiguous, highly creative[1], even in restricted domains. Negotiation or clarification of the input is seen as the only solution to produce more robust, fault-tolerant and user-friendly software integrating natural language processing components.

In this context, the role of the clarification module is to plan a dialogue session that enables the system to recover the missing information the

[1] In the litterature we often read about ill-formed input, but we should probably say unexpected input as what human kind produce has to be called natural language

analysis module has not been able to calculate automatically.

I.2. A proposed architecture

In 'classical' interactive systems, the user's request is built by fusioning the inputs from the different modality channels. The user's request is, then, analyzed, and an unambiguous abstracted request is produced. This unambiguous abstracted request represents, in a way, one of the meanings of the user's request. The unambiguous abstracted request is then transferred (translated) into a sequence of instructions to be performed by the system to answer the user's request. The answer of the system is provided by a generation module, which is in charge of the execution of the sequence of instructions.

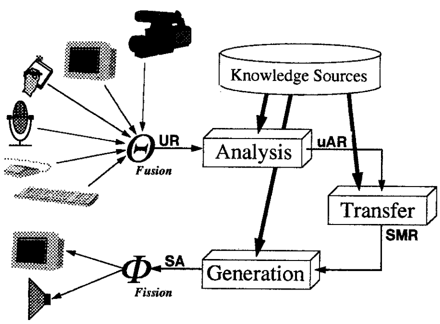


Figure 1: General architecture of interactive systems

The introduction of a clarification module modifies the previous architecture as shown in the next figure.

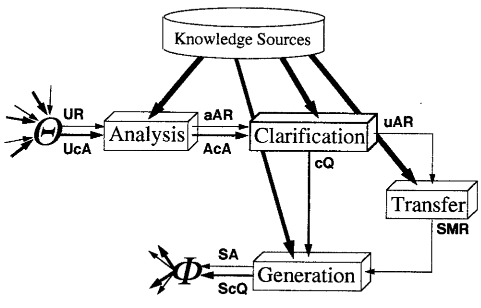


Figure 2: General architecture of interactive systems with a clarification module

UR: User's Request
aAR: ambiguous Abstracted Request
uAR: unambiguous AR
SMR: System Manipulable Request
SA: System's Answer

UcA: User's clarification Answer
AcA: Abstracted clarification Answer
cQ: clarification Questions
ScQ: System clarification Question

With this new architecture, the analysis module produces *all* the solutions corresponding to the knowledge sources used (grammars, dictionaries, knowledge bases, etc.). Thus, the analysis can be ambiguous. If so, the clarification module produces a set of clarification questions that are presented to the user.

I.3. First proposal for a well-formed clarification module

In our opinion the clarification module should be organized around a dialogue planner (or clarification automaton). The dialogue planner is going to use an ambiguity recognition tool (language independant) and an ambiguity knowledge base (language dependent)/

We propose an ambiguity knowledge base made of beams of patterns. In a beam, the patterns describe the differences occurring between the analysis of a sentence when this sentence is ambiguous. Of course, the description of the patterns does not rely on word but on syntactic and semantic local patterns.

To produce the clarification dialogues, each pattern is associated with a dialogue item production method. Such a method describes how the original sentences' components have to be reorganized, reformulated to produce a dialogue item. These methods are described with a basic set of language independent operators.

As far as the content and the presentation of the dialogue are concerned what we can only state now is that this content depends one at least four parameters. The parameters we have identified are the kind of ambiguity; the expertise or needs of the user, the modalities available and the knowledge sources available. The task is now to fill a matrix with one ore several adapted methods.

II. The clarification module in the LIDIA-1.0 mockup

The LIDIA project [Boitet 1989 ; Boitet 1990] aims at studying the concept of Personal Machine Translation, or more precisely, Dialogue-Based Machine Translation for monolingual authors [Boitet & Blanchon 1993], in a multilingual setting

II.1. The LIDIA-1.0 mockup

In the proposed scenario, a monolingual French engineer creates technical documentation, in the form of an HyperCard stack, on a middle-range Macintosh, and helps the system translates it into English, German and Russian. We have opted for a distributed architecture. The author's workstation is a Macintosh and the MT server is on a mini—IBM-4361.

Fig. 3 shows the software architecture of the LIDIA-1.0 mock-up. For it to be easy to read the distributed architecture is not described.

To interact with the system, the user uses the mouse. He can execute actions from menus. He can choose the LIDIA tools on a palette to interact with its document.

Once a textual utterance has to be translated, each sentence of the utterance is first analyzed (MT Analysis) and a multi-solution structure is produced for each of the sentences. The multi-solution structure is a tree structure that reflects all the possible interpretations of the input sentence according to the grammars and dictionaries used by the analyzer. If necessary, clarification questions are produced. When they are answered, the translation process continues off line (MT Transfer & MT Generation). Finally translated documents are produced.

II.2. The clarification module

The clarification module [Blanchon 1994a ; Blanchon 1994b] is made of four elements, two pieces of software: a Disambiguation Automaton and a Pattern Matcher; and two sets of data: a Set of *Patterns Beams* and a set of *Dialogue Item Production Methods.*

The disambiguation automaton organizes the order in which the presence of the ambiguities we have settled is checked. Most of the ambiguities we have settled are defined with one or more beams of patterns. A patterns beam is a set of patterns that share some variables and describe a family of occurrences of an ambiguity. A pattern describes a set of tree.

The next figure shows the patterns beam for an "ambiguity of the construction of the verb – type 3." The beam is made of patterns 8 and 9.

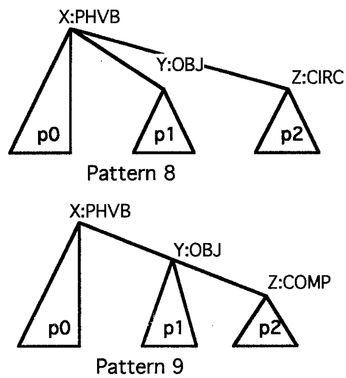


Figure 4: The patterns beam for an ambiguity of the construction of the verb – type 3

A dialogue item production method is associated with each pattern. The methods are described with a set of twelve basic operators. It means that when the previous beam has been recognized, a dialogue with two items will be produced. The methods associated with patterns 8 and 9 are the following.

```
Pattern 8
      Text(?p2), Text(?p0) Text(?p1)
Pattern 9
      Text(?p0) Determiner(?p1)
        Bracket(But_Det(?p1), Text(?p2))
```

If several ambiguities occur in a sentence, a question tree is prepared as shown in Fig. 4.

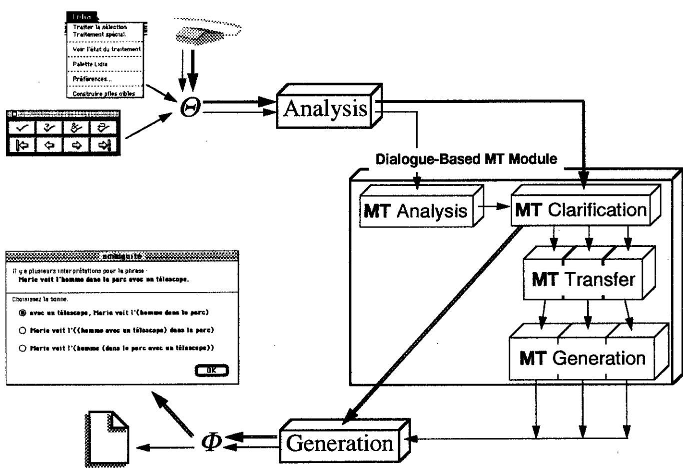


Figure 3: the software architecture of the LIDIA-1.0 mockup

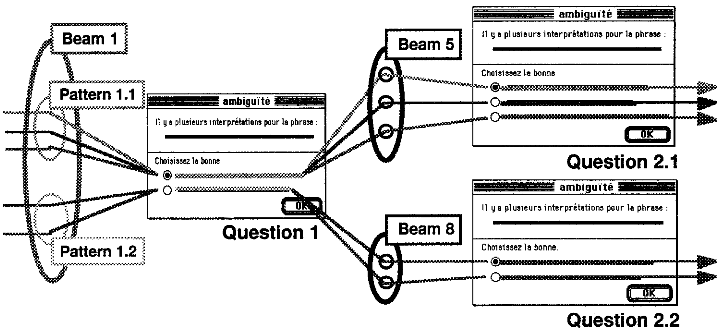
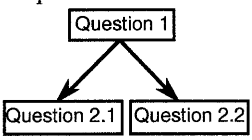


Figure 5: Construction of a clarification tree

The produced question tree is made of the 3 questions:



II.3. An example of a clarification session

For the French sentence "Le capitaine a rapporté un vase de Chine.", the following structure is produced by the analyzer:

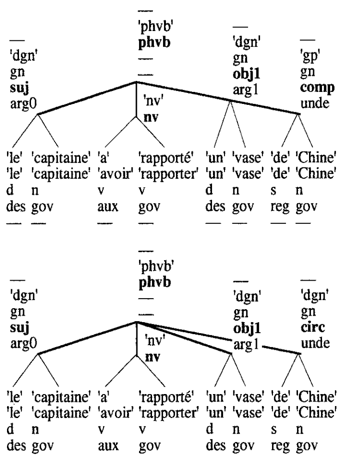


Figure 6: The two solutions produced for the sentence "Le capitaine a rapporté un vase de Chine."

The left solution can be translated in English by "The captain has brought back a Chinese vase." and the right one by "The captain has brought back a vase from China." This ambiguity of this sentence is recognized with the beam of patterns described in Fig. 4. The pattern 9 of the beam matches the left solution with the following matching: **p0**=Le capitaine a rapporté, **p1**=un vase, **p2**=de Chine. The pattern 8 of the beam matches the right solution with, of course, the same matching.

Then, applying the previous described methods, the system will produce the following clarification dialogue.

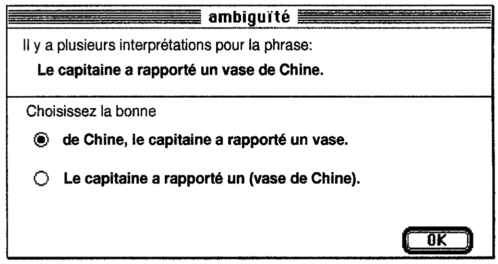


Figure 7: A clarification dialogue [2]

III. Clarification in the framework of MAIPC

III.1. The EMMI setting

For the study of Machine-Aided Inter-Personal Communication ATR-ITL has developed a simulation environment called EMMI (Environment for Multi-Modal Interaction) [Loken-Kim, *et al.* 1993b]. For the time being, with this environment two partners can communicate through one or several wizards of Oz, as shown Fig. 8. In the future, the wizards will be replaced by actual systems.

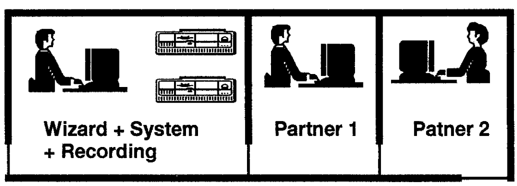


Figure 8: General EMMI setting

2 The proposed interpretations are: ① from China, the capitain brought back a vase, ② The capitain brought back a (Chinese vase).

Each partner is seated in front a workstation (Fig. 9) and can use several modalities to interact with the other. In the foreseen situation, speech is the privileged modality, but a partner can also use a keyboard, a mouse and a touch screen. A video camera is recording each partner's face to be displayed on the other partner's screen. Each partner can hear the other partner and wizard of Oz through headphones.

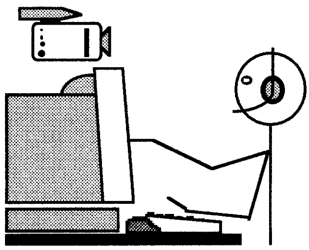


Figure 9: setting of each partner

On the display (Fig. 10), the partners can share a document and interact with it. The document can be for example a map or a form to be filled. There is also a textual window that offers the same services as the Talk process in the Unix environment. Each partner can also display some private document which cannot be interacted with. Finally each partner can control the display of the image of the partner and of himself.

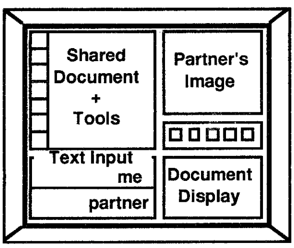


Figure 10: Display of each partner

III.2. Previous experiments with EMMI

The purpose of the first EMMI experiment was to gather baseline data to understand the nature of human-to-human communication in telephone and multimedia settings. Sixteen subjects, eight native speakers of American English and eight native speakers of Japanese took part in the first experiment. They were told to imagine that they had arrived in Kyoto Station, having never been there before, and that they had to find their way to a conference described on a "brochure" they had been given. Their sole means of acquiring this information was by talking to the "conference agent" at the "conference office." The full instructions, as well as transcriptions of the speech data appear in [Loken-Kim, *et al.* 1993a]. Results concerning disfluency, syntactic types of sentences, intention types of utterances, use of deixis and number of turns used, as well as subjects' subjective impressions were analyzed [Fais & Loken-Kim 1994]. An overall finding was that the visual communication options subsumed a significant amount of verbally expressed information. Thus the use of a multimedia environment can reduce the load on a natural language processing system [Fais & Loken-Kim 1994].

The second EMMI experiment continued the investigation of human communication via different communication environments. However, in this second experiment, the subjects, eight English speaking clients, four Japanese speaking agents, and two interpreters, were engaged in a bilingual task. The English-speaking subjects pursued the same tasks as in the first experiment, but in this experiment, communication with the Japanese-speaking agent took place through an interpreter. One change was made to the multimedia configuration; in addition to being able to draw with the mouse, subjects were also able to draw using a touch screen that was added to the system. Subjects were interviewed at the end of the experiment and the auditory data was transcribed as above.

The third EMMI experiment was designed to assess the speech of conversants in a bilingual, multi-modal, machine-translated communication environment. A Wizard-of-Oz design was utilized. Eight English speaking clients conversed with eight Japanese-speaking agents, ostensibly through a machine translation system, in reality mediated by two interpreters, one a native speaker of English translating from Japanese to English, the other a native speaker of Japanese translating from English to Japanese. All subjects were interviewed at the end of the experiment.

Disfluency measures as well as other linguistic parameters will be analyzed for both the second and third EMMI experiments. The speech data from the third experiment will be analyzed to determine how subjects modify their speech in response to a failure on the part of the "machine" to understand them. In addition, all three experiments will be examined for lexical accommodation and simultaneous speech phenomena.

III.3. Towards clarification experimentation within the EMMI setup

The clarification module developed for the LIDIA-1.0 mockup is being reprogrammed with better software engineering techniques following the guideline we proposed in § I.3. Figure 11 shows its organization.

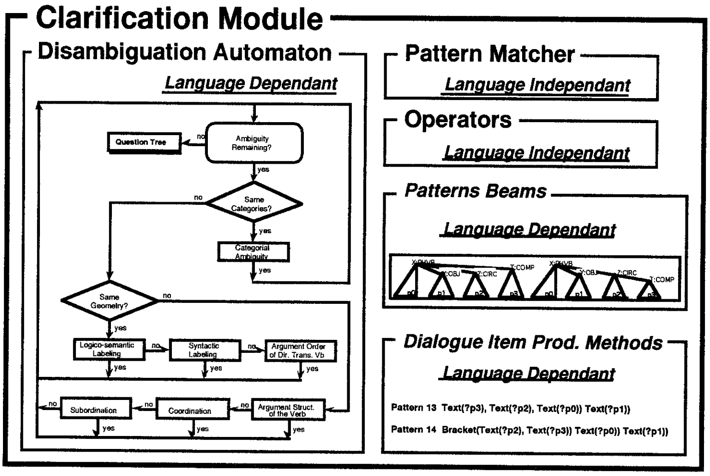


Figure 11: Architecture of the clarification module

We are also adapting it to English input. To complete this task, the first step has been to gather a corpus of ambiguous sentences from the data collected with EMMI. The second step was the realization of the parse trees. Up to now the trees are made by hand. We were then able to classify the ambiguities. We will now define the beams of patterns and the dialogue item production methods.

With the data collected in the framework of the third EMMI experiment we will be able to see if the wizards have used clarification questions, in what proportion and what the questions were about. This will give us a first set of information. We also consider to propose new experiments to study the relevance of new clarification scenarii in a multi-modal setting.

Conclusion

We have tried to justify the need for clarification module in the context of natural language interface and MAIPC.

The clarification module we have developed in the framework of BDMT at the GETA Laboratory in France and its adaptation to English input at ATR-Interpreting Research Laboratories in Japan allowed us to sketch a first proposal for well-formed clarification modules.

What do we expect from a common agreed framework is the possibility to exchange and compare data and methods. In such framework tools like the EMMI setting are a very important to develop and test clarification methods.

There are also a lot of questions we did not tackle yet as: – the need for the clarification or the parsing modules to learn about the user, – the need to calculate some weight for each interpretation so as not ask question is some threshold is reached, – the need for the clarification module to build description of new ambiguities using reasoning techniques.

References

Blanchon H. (1994a). *Pattern-based approach to interactive disambiguation: first definition and implementation.* Rap. ATR-Interpreting Telecommunication Research Laboratories. Technical Report. n° TR-IT-0073. 8 Sept. 1994. 91 p.

Blanchon H. (1994b). *Perspectives of DBMT for monolingual authors on the basis of LIDIA-1, an implemented mock-up.* Proc. Coling-94. Kyoto, Japon. 5-9 août 1994, vol. **1/2** : pp. 115-119.

Boitet C. (1989). *Motivations, aims and architecture of the Lidia project.* Proc. MT SUUMMIT II. Munich. 16-18 août 1989, vol. **1/1** : pp. 53-57.

Boitet C. (1990). *Towards Personnal MT : general design, dialogue structure, potential role of speech.* Proc. Coling-90. Helsinki. 20-25 Août 1990, vol. **3/3** : pp. 30-35.

Boitet C. & Blanchon H. (1993). *Dialogue-based MT for monolingual authors and the LIDIA project.* Proc. NLPRS'93. Fukuoka, Japon. 6-7 décembre 1993, vol. **1/1** : pp. 208-222.

Caelen J. (1994). *Multimodal Human-Computer Interaction.* in Synthesis and Speech Recognition. Keller, E. (ed.). John Willey & Sons. New-York. 40 p.

Fais L. & Loken-Kim K.-H. (1994). *Effects of communicative mode on spontaneous English speech.* Rap. Institute of Electronics, Information and Communication Engineers. Technical Report. n° NCL94-22. Oct., 1994. 8 p.

Goddeau D., Brill E., Glass J., Pao C., Philips M., Polifroni J., Seneff S. & Zue V. (1994). *GALAXY: a Human-Language Interface to On-Line Travel Information.* Proc. ICSLP 94. Yokohama, Japan. September 18-22, 1994, vol. **2/4** : pp. 707-710.

Haddock N. J. (1992). *Multimodal Database Query.* Proc. Coling-92. Nantes, France. 23-28 juillet 1992, vol. **4/4** : pp. 1274-1278.

Hiyoshi M. & Shimazu H. (1994). *Drawing Pictures with Natural Language and Direct Manipulation.* Proc. Coling-94. Kyoto, Japon. August 5-9, 1994, vol. **2/2** : pp. 722-726.

Kay M., Gawron J. M. & Norvig P. (1994). *Verbmobil: A Translation System for Face-to-Face Dialog. CSLI lecture note no 33.* Center for the Study of Language and Information, Stanford, CA. 235 p.

Loken-Kim K.-H., Yato F., Fais L., Kurihara K., Furukawa R. & Kitagawa Y. (1993a). *Transcription of spontaneous speech collected using a multi-modal simulator — EMMI, in a direction-finding task (Japanese-Japanese; English-English).* Rap. ATR Interpreting Telecommunication Research Laboratories. Technical Report. n° TR–IT–0029. Dec., 1993. 101 p.

Loken-Kim K.-H., Yato F., Kurihara K., Fais L. & Furukawa R. (1993b). *EMMI – ATR environment for multi-modal interactions.* Rap. ATR Interpreting Telecommunication Research Laboratories. Technical Report. n° TR–IT–0018. Sept. 30, 1993. 28 p.

Morimoto T., Suzuki M., Takezawa T., Kikui G., Nagata M. & Tomokio M. (1992). *A Spoken Language Translation System: SL-TRANS2.* Proc. Coling-92. Nantes, France. 23-28 juillet 1992, vol. **3/4** : pp. 1048-1052.

Nishimoto T., Shida N., Kobayashi T. & Shirai K. (1994). *Multimodal Drawing Tool Using Speech, Mouse and Key-Board.* Proc. ICSLP 94. Yokohama, Japan. September 18-22, 1994, vol. **3/4** : pp. 1287-1290.