Internal Use Only

TR-I-0017

Summaries of Workshop on Natural Language Dialogue Interpretation

自然言語対話理解ワークショップ講演要録

November 27-28, 1987
ATR Interpreting Telephony Research Laboratories

December, 1987

ATR Interpreting Telephony Research Laboratories
ATR 自動翻訳電話研究所

Foreword

Research, both theoretical and practical, into computational treatment of natural language dialogue is now widespread. Our automatic telephone interpretation project at the ATR Interpreting Telephony Research Laboratories is, we hope, a contribution to a far larger overall enterprise.

To bring together some of the many researchers in the field, we held a small invitational workshop on "*Natural Language Dialogue Interpretation*" here in Osaka, on November 27 and 28, 1987. With a great contribution from outstanding speakers, we had a fruitful and successful workshop.

The program included presentations centered around computational treatment of machine translation for natural language dialogue such as Theoretical Considerations of Dialogue Understanding, Applications of Knowledge to Analysis and Translation of Dialogues, Framework for Dialogue Translation, Problems Involved in Telephone Interpretation, Technical Prospects of Telephone Translation, etc..

This report consists of the extended abstracts of the papers presented at the workshop. Some of the abstracts were transcribed and summarized by ATR members, not by the authors themselves.

Table of Contents

Opening Address

Kohei Habara
Advanced Telecommunications Research Institute International

It is a great pleasure for me to welcome all of you to the first ATR Workshop on "Natural Language Dialogue Interpretation". I would like to thank each and every participant for their interest and efforts in helping us make this workshop possible, especially those who have traveled great distances and taken valuable time from their very busy schedules. We are particularly fortunate in having with us five scientists from the U.S.A.

For those of you who are not familiar with ATR, I would like to present a brief background on who we are and what we are doing.

ATR, or more precisely the Advanced Telecommunications Research Institute International, was established in spring of 1986 with support from industrial, academic, and governmental organizations to serve as a major center for basic and advanced telecommunications R&D.

Precisely speaking, "ATR" is the generic name for five organizations, namely ATR International and four R&D corporations supported by ATR International. The four R&D corporations are:

the ATR Communication Systems Research Laboratories,
the ATR Interpreting Telephony Research Laboratories,
the ATR Auditory and Visual Perception Research Laboratories, and
the ATR Optical and Radio Communications Research Laboratories.

70% of the capital of each of the R&D corporations is funded by the *Japan Key Technology Center,* and 30% by private enterprise. The Japan Key Technology Center is a governmental organization established in October 1985 as a central institution to promote basic technological research in the private sector. The budget of this center is obtained from dividends that the Japanese government gets from Nippom Telegraph & Telephone Corporation, or NTT, after its privatization. The Japan Key Technology Center promotes basic research of key technologies by investing its capital in newly-established R&D corporations corresponding to each new project. About 30 projects have been started, the largest four of which are ATR in-house projects. That is the reason why we have 4 R&D corporations. For that reason, organizational structure of ATR is somewhat complicated. Essentially, ATR is single organization indeed.

Our research institute has been provisionally housed in this building, located in one of the most beautiful places in Osaka, since April, 1986. However, in 1989, ATR will move to its new research facilities now being constructed in the Kansai Science City. This move will insure an appropriate, conductive environment for basic and creative research. And it will also suit the central role we intend to play in the Science City.

Let me turn now to the research goals of our four R&D corporations. As a participant of this workshop, you are probably most interested in the ATR Interpreting Telephony Research Laboratories, but before I introduce them, let me quickly mention the other three.

The goal of the ATR Communication Systems Research Laboratories is to develop communication systems based on human needs and designed for ease of operation. These researchers are devoting themselves to basic research in such areas as full-sensory communications, nonlinguistic communications, high-security networks, and the automated development of communication software.

The goals of the ATR Auditory and Visual Perception Research Laboratories are to make technological breakthroughs in the fields of visual pattern and speech recognition, and to develop efficient and friendly man-machine interfaces. The basic mechanisms of perception and cognition in the human senses of sight and hearing are now being investigated from engineering, psychological, and phisiological standpoints.

The ultimate objective of the ATR Optical and Radio Communications Research Laboratories is to establish technologies that will allow anyone, at any time and at any place to communicate. From this premise they are conducting basic research in optical and radio communications focusing on space and mobile communications and on communications devices based on artificially modulated material structures.

Finally the goal of the ATR Interpreting Telephony Research Laboratories is motivated by the vision that one day, through technological advances, the people of the world might be able to communicate in their own language with each other smoothly and comfortably. Such an automatically translating system requires three major components:

Speech recognition to identify the spoken message in one language;
Machine translation to translate this message from one language into another; and
Speech systhesis to automatically generate natural sounding human

speech in the target language on the other end.

The considerable basic research required by each of these component technologies is now being carried out, in an effort to achieve this ultimate in telephone communication.

This workshop was organized by the ATR Interpreting Telephony Research Laboratories as a forum to focus on the problems involved in the computational treatment of telephone conversations and to improve our understanding of human dialogues.

In my feeling, spoken languages are somewhat different from written languages, especially in Japanese, and studies on the spoken languages have just started. I am sure that this workshop will provide new insights into many of these problems. And, I hope that this workshop would be commemorated as a milestone in the study history of spoken languages in the future.

I also hope that this workshop will provide the opportunity for personally exchanging scientific results, that it will facilitate the making of new acquaintances, and I sincerely hope that it will prove effective in the strengthening of friendships among the participants from different parts of Japan and the U.S.A.

Finally, I hope that some day in the future, I may give this address in Japanese, thanks to the development of our Interpreting Technology.

KEYNOTE SPEECH

PROBLEMS INVOLVED IN TELEPHONE INTERPRETATION

Akira Kurematsu

ATR Interpreting Telephony Research Laboratories
TWIN 21 Bldg. MID Tower
2-1-61 Shiromi Higashi-ku Osaka 540 JAPAN

1. Introduction

The Automatic Telephone Interpretation system is a facility which enables a person speaking in one language to communicate readily by telephone with someone speaking another. It does so by automatically and simultaneously transforming the dialog from the speaker's language to the listener's. Three constituent technologies are necessary for such a system: speech recognition, machine translation, and speech synthesis.

Since this is a brand-new concept, numerous studies and evaluations on applicability must still be made. A high degree of performance capability from each of the constituent technologies and system "user friendliness" of the system are also essential. According to a feasibility study sponsored by the Japanese Ministry of Posts and Telecommunication, perfecting a system like this will require fifteen years.

2. Research Tasks and Approach

In the areas of speech recognition, machine translation and speech synthesis, some technological progress has already been made. This progress, however, this is still far from the level needed for a viable system. Thus, extensive fundamental research is still necessary in each of these fields.

2.1 Speech Recognition

(1) Recognition of continuous speech

Since conversational speech is normally continuous, with most words running together, recognition of phrases of continuous speech is necessary. This is especially true in Japanese, where stem boundaries, verb endings and adverb endings are not explicit and humans speak language without much awareness of the definitions of postpositional particles or auxiliary verbs in conversational utterance.

From a linguistic viewpoint, speech can be broken down into minimal units known as "phonemes." For recognition of continuous speech, the phonemes must be discerned first, then can words and phrases be recognized. The main approaches to phoneme recognition either employ features base or use the Hidden Markov model or neural net model.

(2) **Large vocabulary speech recognition**

The automatic telephone interpretation system must eventually recognize vocabularies ranging from a few thousand to several tens of thousands of words. The ultimate aim of the ATR project is a technology that will recognize continuous speech containing a vocabulary of approximately 3000 words with several hundred specific task words. Computation increases drastically, however, with expanded vocabulary. An algorithm based on word spotting will solve this problem.

(3) **Speaker adaptation**

One of the most problematic aspects of speech recognition is that speech characteristics differ from one speaker to another. If the interpretation system is to be made available for general use of various ages, speaker-independent speech recognition technology will have to be developed.

The most effective approach to this problem will be the incorporation of a system for speaker adaptation. The role of this system is to recognize the speech characteristics peculiar to a given speaker, so that the necessary adaptations may be made. Research is needed into algorithms that will enable speaker adaptation based on a limited amount of speech data and information on various possible speech variations already in the system. As a means of spectral pattern learning, one promising approach is the use of vector quantization for speaker adaptation.

(4) **Speech database**

A speech database is essential if the above research is to be carried out efficiently. Speech data will be gathered in the form of important words and continuous short sentences. Then labels, including the name of the phoneme and other important information, based on the spectrograms will be attached. Conversational speech data on some domain topics will also be accumulated.

(5) **Understanding of conversational speech**

Conversational speech has a structure which differs greatly from single utterances or recitation. Therefore, it will be necessary to study the characteristics of utterance and the grammatical structure peculiar to conversation. Moreover, the treatment of meaningless throat-clearing, interjections and other sounds will also be necessary.

A technology which uses language information and background knowledge to understand speech content is necessary. Knowledge processing or language processing will also be approached. Conversational speech recognition in some domain will be considered. Studies of syntax, semantics and topical analysis, as well as creation of a model for speech understanding will be developed.

(6) Prosodic and linguistic processing

Prosodic information such as pitch, stress, and duration, along with information on syllable boundaries, will be used in order to increase the precision and speed of algorithms for phoneme recognition and word spotting. Analysis will also be made of the correlation between prosodic and linguistic information.

(7) Influence of interference

Telephone speech is mixed with surrounding noises and telephone circuitry interference. Establishment of a technology for speech recognition which eliminates these influences is vital. To overcome such interference, a signal processing technology that will eliminate interference from sound will be studied.

2.2 Machine Translation

The goal of machine translation research in this project is to establish a machine translation technology for Japanese/English conversation.

Systems presently available for translating written texts from one language to another are applicable only to certain limited fields. Moreover, they require human intervention both for pre-editing the text to be translated, and for post-editing the resulting translation. Further development of semantic processing techniques, and of the ability to consider context, is required in order to raise machine translations to an acceptable level. Spoken language differs from ordinary written language in both vocabulary and grammar. The analysis of a number of linguistic phenomena and research into computer implementation schemes are necessary.

Considering the growing use of electronic mail and other text telecommunications, the necessity of translating these is evident. Furthermore, foreseeable trends in the development of new telecommunications media would seem to indicate that an automatic telephone interpretation system should not be limited to speech. The most efficient approach would be to develop a system that is also applicable to mixed media containing text and speech information.

Correspondence differs from technical documents or articles. They contain expressions not ordinarily found in written language. For this reason, research into sentence patterns and contextual matters characteristic of correspondence will be needed, so that a translation method may be developed for document telecommunications. It is also felt that this research will be provide basic data for conversational spoken language translation.

(1) Dialog structure

Dialog model research as well as a broad and thorough explanation of the dialog process will be necessary. Studies focusing on dialogs that take place via telephone or keyboard will be carried out. Research will be done on linguistic models that enable dialog generation. Research on discourse structure will include examining the conditions making dialog possible.

One area to be emphasized is research on techniques for recognizing dialog structural segments. The following items will also be studied in Japanese; analysis of the role connectives and connective-like prepositional phrases at the beginning of statements; analysis of certain styles of indicating the subject; establishment of conversation units; recognition of the focus, theme, and subject, as well as clarification of their relative to the dialog; and the relationship between the segmental structure of a dialog and ellipsis / anaphora.

(2) Rules governing speech dialog and dialog processing system

An effort must be made to formulate rules for the linguistic presuppositions, or implicit conditions, that facilitate the development of a dialog. Rules will be formulated to define the framework for the application and inheritance of these presuppositions. Work will also be directed toward verifying and extending conversational postulates for dialogs in general. Studies will be carried out to systematize and formulate rules for the implicit conditions underlying conversation. Investigations of the rules governing the use of pronouns, demonstrative pronouns, and ellipses will also be undertaken.

Studies will be carried out to find a processing mechanism that can understand conversation units as a dialog proceeds. Research will primarily focus on identifying anaphora, confirming the contents of ellipses, clarifying the process by which linguistic presuppositions are inherited and transmitted, and predicting changes in circumstances using associational knowledge.

(3) Schemes for understanding meaning in context

It is essential that the telephone interpretation system be able to comprehend the meaning in context. Investigations will be made into schemes for understanding the meaning of each conversation unit, taking into consideration the conditions surrounding the development of a dialog. The main areas of research will be: how to accurately represent the speakers' state of comprehension, which changes according to the surrounding circumstances, how to predict what will be said next, and how to manage changes in the speakers topics and statements.

(4) Inference mechanism designed for understanding dialog

The process of understanding a dialog involves an ongoing construction of the state of comprehension. To do this, it is necessary to generate and modify the surrounding circumstances through both knowledge contained within, and existing outside of the language.

Studies will be done on methods of applying rules of inference, high-speed unification mechanisms, and ways of using associational knowledge.

(5) Lexical databases and knowledge bases

A lexical database containing the terminology occurring in conversational sentences and correspondence texts, as well as the idiomatic expressions that appear frequently in conversation, is a requirement for realizing a high-level machine translation system. Also essential is a large-scale database that includes grammatical rules.

Translation ambiguity will be reduced only when a deeper understanding of the meaning of words is achieved. If the domain or task of a conversation or correspondence is specified, then possessing specialized knowledge of that domain is obviously a help in understanding the content. Technology must therefore be established for grasping and describing that knowledge conceptually, along with the mutual relationship of concepts. The concepts expressed by words, and the explanations of those concepts, will need to be systematically organized into a knowledge base. They can greatly afford assistance in the process of understandingmeaning.

(6) Knowledge representation schemes

Studies must be carried out to find a suitable scheme for representing linguistic knowledge as well as common sense outside language and specialized knowledge peculiar to a specific domain. The results obtained will be used to establish a flexible knowledge representation scheme suitable for natural language understanding.

(7) Interface of speech recognition and language processing

Perfect speech recognition cannot be expected. Information from the speech recognizer to the language processor will be in the form of phrase lattice. It will be necessary to select the most appropriate candidate or to enhance the error correction system so that it can comprehend spoken dialog.

In the language processing systems, it will be necessary to have a function that can use syntactic and semantic knowledge to correct errors in the input sentences. The knowledge that is to be used will be rules of syntax, semantics and pragmatics.

2.3 Speech Synthesis

Speech output produced at the end of the automatic telephone interpretation process will be based on techniques for converting texts to synthetic speech. Both clarity and naturalness of synthesized speech must be improved.

(1) Highly natural synthetic speech

The main factors in speech clarity and naturalness are the proper selection of the speech synthesis unit and control of the rules of prosody. An effort to use compound speech synthesis units of various lengths will be made. Rules will have to be established regarding the deformation and connection of speech synthesis units when they are combined to form the words and phrases that make up continuous utterances.

Another matter for research is prosody, which relates closely to meaning and naturalness. When the resulting translation is converted to speech, that speech should have a "tone of voice" in keeping with the meaning of the utterance. This involves finding ways to control such factors as accent, stress, and speed on the basis of meaning, and will require further research into the incorporation of conceptual information in speech synthesis.

(2) Individuality of voice quality

The synthesized speech should resemble the actual speaker's voice as much as possible. It should also be possible to synthesize the different speech qualities of, for example, younger and older people. The required techniques include extraction of the factors involved in speech quality information, and methods for controlling speech quality.

3. Conclusion

In closing, let me reemphasize that three areas of speech recognition, machine translation, and speech synthesis are necessary to realize a system for automatic interpretation of telephone conversations.

An attempt will be made to raise the level of technology in each of these areas qualitatively, based on enormous volumes of linguistic data. Since speech processing and machine translation deals with natural language, it will necessitate developing techniques for constructing massive databases of speech and language information.

To implement real-time automatic telephone interpretation will require advanced technology of high-speed processing. Moreover, because of the diversity of parallel processing required, techniques for creating software systems will need to be investigated.

The application of the configuration of an automatic telephone interpretation system to a telecommunications network will also to be considered. From the standpoint of human users, research in the communications science field is also required for solving the human problems involved in man-machine interaction.

Research into automatic telephone interpretation has the primary aim of overcoming the language barrier to international telecommunications. Because of the vast complexity of natural language, however, this goal can only be reached by gradually improving levels and techniques. It is important, in other words, to proceed with an insight into what is, and is not, technically feasible.

One more important requirement of this ambitious project is the international cooperation of research institutions in Japan and abroad.

Distinctive features of Conversations and Inter-Keyboard Interpretation

Hitoshi IIDA

ATR Interpreting Telephony Research Laboratories

Twin21 Bldg. MID Tower, 2-1-61 Shiromi, Higashi-ku, Osaka 540, Japan

1. Introduction

This talk demonstrates that the distinctive features of telephone and keyboard dialogues are clarified on the basis of simulation data analysis. Keyboard dialogue has been shown to have the same fundamental conversational features as telephone dialogues except for stammering, correction, and interjections indicating agreeable responses. Both types of dialogue have similar ways of expressing, dialogue characteristics, and discourse structuring in terms of the 'discourse segment'. We feel that the results of our research justifies a study of keyboard dialogue as an approximation of telephone dialogue. And I will also introduce a new approach to keyboard interpretation.

2. Inter-Keyboard Dialogues

Our target domain is, of course, telephone dialogue. Other methods of communication used to exchange information, such as 'telex' and 'email', are also of interest to us. These are very similar to texts ; however their grammars and vocabularies depend on special domains. They are one-way communication and not a duplex or half-duplex transmission. Their merits are that it's possible to keep a written dialogue history, so can easily be referred to former descriptions. On the other hand, telephone dialogues are placed in speech dialogues and bi-directional transmission. These characteristics will have to be handled by a future interpreting system because they are the foundamental elements of spoken dialogues. Presently, telephones have no monitoring display equipment and some expressions, especially references, are apt to be mistaken. There is no need to worry about these problems with inter-keyboard dialogues. We have just got a new research domain, inter-keyboard dialogues.

3. Differences between Telephone Dialogues and Inter-Keyboard Dialogues

3.1 Cohen's comparison

Before going into ATR's comparison of telephone dialogues (TD) and our new subject inter-keyboard dialogues (KD), let's consider a study done by Cohen. He compared TD with KD using frequency distribution of 'communicative acts', which are illocutionary acts, for example, request, promise, inform and so on. On the other hand, our comparison of TD and KD concerns 'dialogue expressions' usage.

He set a task of assembling a toy water pump. The communication situation was between an expert and an apprentice. In the task, the expert directed the apprentice to assemble a water pump by using a telephone and keyboard.

One typical result of his comparison showed that one specific kind of communicative act dominates in an instruction-giving dialogue. That act is a request that the hearer idetifies the referent of a noun-phrase. That request is nearly absent in keyboard dialogues. These requests are only achieved "indirectly".

3.2 ATR's comparison

We set a task centered on inquiries regarding attendance at the First International Conference on Interpreting Telephony, and prepared a call for papers of which is given to a secretary for the conference and an inquirer. The instructions given to them were to use the call for papers and common sense to make inquiries about attendance. Such a dialogue has a task goal much like the dialogue about assembling a pump had. However, there are some diferences. In our attendance dialogue a turn or change in speaking initiative is evident.

An intention type set is established and each type is placed on a surface intention. The first example is "o-namae-wo o-kika se negaemasu ka.," that is 'please, your name.' as a very simple meaning. The phrase "negaemasu ka" placed at the end of the utterance has the intention of demand in a polite request form.

The second example is "kaigi-ni sankasi tai nodesu ga.," that is 'I would like to attend the conference.' The last word "ga" is a conjunction, so "ga" follows a subordinate clause and after it the main clause follows. But, in this case, the main part is not described and the subordinate part plus "ga" expresses an euphemistic communicative act. They are used to so often the intentions rather than to actually form subordinate clauses.

Main characteristics are using "utterance delimiter", "ellipsis", "idiom-like phrase indicating intentions", and "clue words or phrases". The interjections, "ee", "a", "ano" and the like appeared in TD, but not in KD. The response signal, "hai," signifying agreement, appeared most frequently in TD, but not in KD. The words "soodesuka" (is that so?) and "wakarimasita" (I understand), signifying that what the partner had said had been read or heard and fully understood were seen in both types of dialogues.

Differences in ellipsis usage is found in both types of dialogue in the way postpositions and topics were omitted. In the telephone data, there are a number of case postposition omisions, such as "ga" (nominative case marker), "wo" (accusative case marker), and "ni" (dative case marker). These are seldom omitted in KD.

Ways of expressing intention, which is characteristc of spoken Japanese dialogue, is seen in both TD and KD. The conjunction "ga" in the example,

"hai, deki masu ga," which means "Yes, we can.,"

is not needed logically. "Ga" is used only for dialogue conventions and it appears both in TD and KD.

The relationships between the constituent factors of TD and KD are illustrated on the figure. Factors such as situation dependent meaning, intention expressions by means of idiom-like phrases, ending signals, clue words and phrases, discourse structures, and some of the response signals, repetition expressions, and the elliptical expressions are found both in TD and KD.

Interjections, corrections, substitutions, stammering and so on can be regarded as noise in natural language processing. Because KD are word sequences without noise whose main dialogue characteristics are maintained, and whose main elements are intentions using conventional expressions it is quite reasonable to assume that they can be effective used in natural language dialogue understanding applications.

4. Problems on KD Interpretation

There are at least two main and important problems. One is how to maintain dialogue smoothness. The other is how to keep a dialogue in one language correspondent to a dialogue in the other language.

The following expressions are mainly used to keep a dialogue smooth.

ⓐ idiom-like phrases indicating intentions
ⓑ euphemistic or polite forms
ⓒ ellipsis (omitting old informations, ie. topics)
ⓓ references
ⓔ utterance segmentation signals
ⓕ clue words

The expressions from ⓐ to ⓒ define inter-sentential meanings in the local region of a dialogue. The expressions from ⓓ to ⓕ are concerned in the entire dialogue's meaning and discourse structures. Idiom-like phrases indicating intentions are, of course, determined precisely by means of context.

5. Understanding inter-sentential meanings

Full use of the characteristics of each word must be made for analysis, translation, and generation. We are taking the approach of constructing and decomposing feature structures by means of calculating feature structures under a unification-based approach. Handling dialogues by means of feature-based descriptions will give us the following advantages over other methods.

It is easier to handle Japanese phrases indicating intentions.
It is easier to understand such a form.
It makes it possible to transmit a topics to the next utterance.

We discuss a feature structure of one utterance,

"sankaryou-ha iru no deshou ka" which means "Is an attendance fee

necessary?,"
"sankaryou," that is an attendance fee, is marked by the particle "wa," topicalized and moved to the top of the utterance. The phrase "no desho u ka," which indicates some intentions, consists of some auxiliary verbs and a particle, and is complexly connected. Then we introduce a special feature, "topic", and we can handle a topicalized phrase under one utterance and also under a structure upper the utterances. Such description systems as JPSG can transmit topic features to the next utterance structure

Next, a process of translating an example is briefly shown. The example is "kaigi-ni sankashi tai nodesu ga" which means "I would like to attend the conference." Of course, it lacks the subject, but it is obvious that it's the speaker. And there appears a phrase indicating some intentions, that is "nodesu ga". This phrase can be devided into two meanings. "nodesu" indicates one communicative act which is "INFORM for EXPLANATION" and the other word, conjunction "ga" indicates '+' euphemistic value.

A semantic structure of an example has propositional content part, intention part, expression type and so on. The propositional content part is described in the form of Head-driven Grammar-like and consits of two features, content and indices. Content value is basically a case structure and in this content there are two relation properties. And Indices value is a variable set. Intention part is a communicative act, INFORM-EXPLANATION, and euphemistic value is plus. That's a semantic representation of the sample utterance.

In translation process a propositional part in Japanese must be only transfered into one in English. The process basically transfers a case structure from Japanese to English. And the generation process decides a suitable expression for the intention-part and other properties. It is possible to decompose intentions into a suitable surface linguistic expression by means of unification calculus.

Example-1 (keyboard conversation)

A: How much is the registration fee ?
B: A hundred dollars U.S.
A: Can this be paid in yen ?
B: Sorry. We must request payment in dollars.
A: Is it 100 dollars for all three days ?
B: Yes. Hotel room charges and meals are not included.
A: I understand.

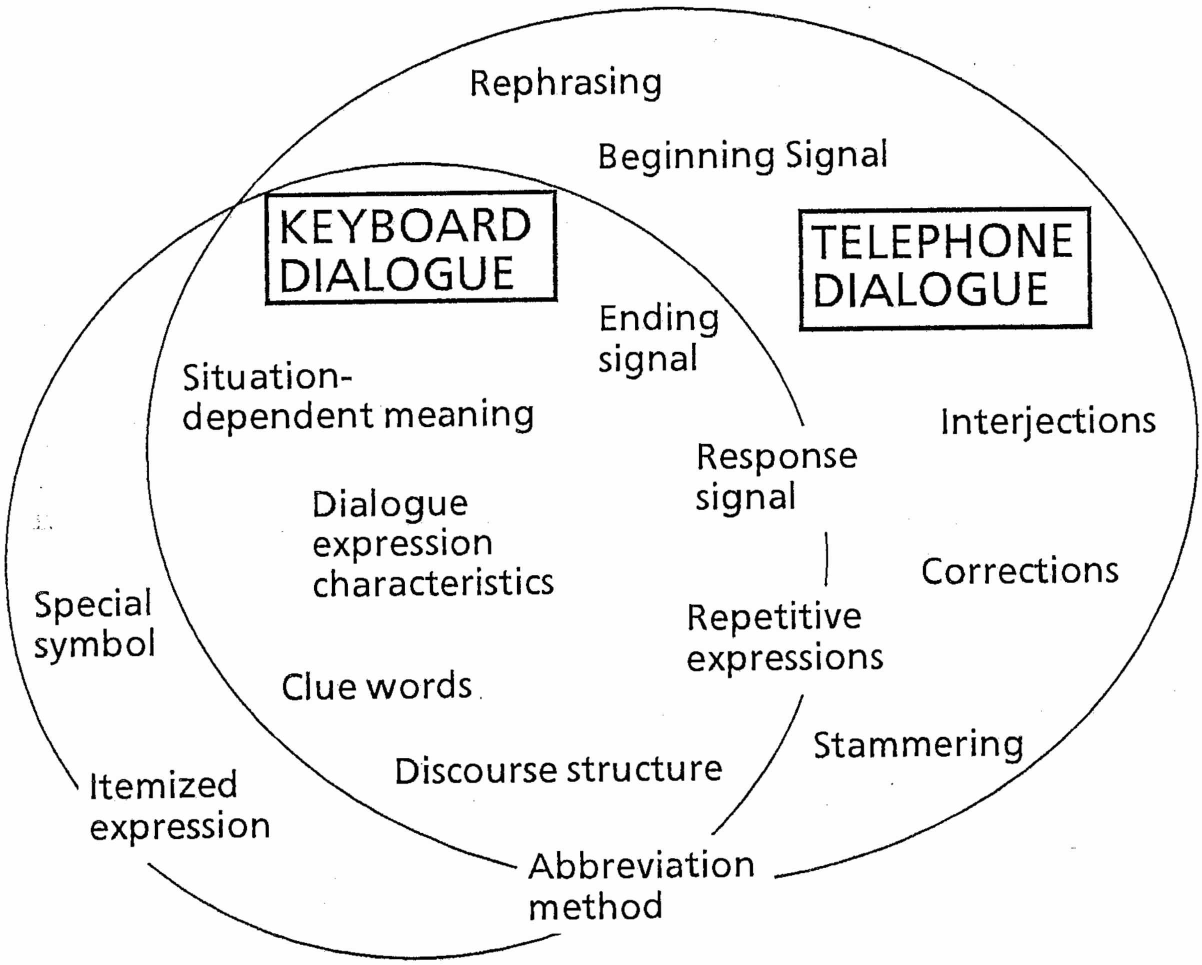


Figure-1 Implications of Telephone and Keyboard Dialogues

Dialogue translation vs text translation:Interpretation-Based Approach

Jun-ichi Tsujii (Kyoto University)

Introduction

After experience with Machine Translation (MT) of texts, dialogue translation (DT) and text translation (TT) are both difficult, but in very different ways. Discussing the differences of DT and TT systems, the conclusion was reached that certain types of the dialogue make a DT system more feasible than a TT system.

Different environments

Among various MT engineering problems to be considered, the below question should be considered from the viewpoint that different application environments might require different technologies.

How can a feasible MT system which can be used in an actual translation environment be designed?

To answer this question concretely, two further questions should be answered.

Questions:
(1) In actual application environments, which is more feasible, a DT system or a TT system?
(2) Can a capable DT system be designed simply by modifying current MT technology for TT?

Answer for (1)

A certain type of DT system *may be more feasible* than a TT system,

---

because of the basic difference in the environment where these systems might be used.

(a) *Clear Definition of information*
In certain types of dialogue sentences (goal-oriented types such as hotel reservation, conference registration, doctor-patient dialogue), *what information should be transmitted from source sentences to target translation* can be defined more clearly. This is not true of text sentences.

(b) *Active Participation*
In DT systems, there are real participants -*Speakers* and *Hearers*- who really want to communicate information at the time of translation, while TT systems are supposed to be used by professional translators, which means no one is really involved in the text at the time of translation.

These differences might make a DT system more feasible, if the system is properly designed to take advantage of them.

Answer for (2)

System organization of a DT system *should be quite different* from that of a TT system,

---

in order to take full advantage of the DT environment, system organization should be different from that of TT. Mere extension of current MT technology to TT will not result in high quality DT systems.

Dialogue Translation Systems

In a simplified framework of natural language understanding application (Fig. 1), understanding a sentence is regarded as a process of tranformation from an input sentence into a *meaning representation* of the sentence. In this framework, information to be extracted from the input is pre-defined depending on aims of the internal task. In other words, in this schema, *meaning representation* is a kind of extracted result relevant to the specific process.

However, the process of translation itself does not define what sort of information is to be extracted because the aim of translation is re-expressing in sentences of the target language *all aspects of the information contained in the source sentences,* with the least possible distortion.

For texts, available information can not be distinguished from all aspects of information. As a consequence, most actual TT systems make attempt to extract structural descriptions of source sentences and their translation results are inherently *structure bound.* However, in certain types of DT, we can define important information by refering to the aim of dialogue.

A schematic view of a hypothetical system (Fig. 2) divides utterances into parts containing important and less important information. The system can treat these differences differently in translation (Fig.3(a)). A *layer of explicit understanding* contains interpretation of information which can be represented in a language-independent, but task-dependent, way. For less important information, it is permissible to get structural descriptions through actual TT technology. A translation system which extracts important information from source utterances can produce *less structure bound* translations. Varieties of surface expressions can be reduced to a single meaning representation, if they convey essentially the same information (Fig.3(b)). English translated expressions will be chosen on the basis of English context rather than Japanese structure. As an extreme, a system which produces fluent translations only for important parts of utterances (Fig.3(c)) is imaginable.

To introduce this *explicit understanding layer*, knowledge about the specific domain, its natural flow of dialogue, and its surface linguistic expressions should be system prerequisite. By utilizing this layer, we can expect *less structure bound* translations of the important parts of utterances.

In general, MT systems need not be able to *understand* completely the input texts. The human reader may understand the translations, if important information is translated properly. Moreover, in dialogue situations, if the outputs are difficult to understand, the hearers can ask quesions to clarify the speakers' intention.

Furthermore, the system can pose questions to the speakers if it cannot capture or *understand* the important parts of utterances. The existence of real participants indicates the possibility of extending the role of the translation system to that of *coordinator of communication.*

Conclusions

A DT system is more feasible than TT system if properly designed. In order to develop a feasible DT system, the following key points should be considered;

[1] Different levels of *descriptions*
By treating the two layers - the explicit understanding layer and the structural description layer - differently, at least the important information can be processed throught the DT system.

[2] *Flexible interactions* during the translation process
The existence of dialogue participants at the time of translation can promote the avilability of a DT system.

[3] Management of *dialogue structures*
To find important information in discourse, a DT system should be able to utilize various knowledge such as knowledge of surface clue expressions, task dependent knowledge, knowledge of discourse structures, etc.

(summarized by M.Kume)

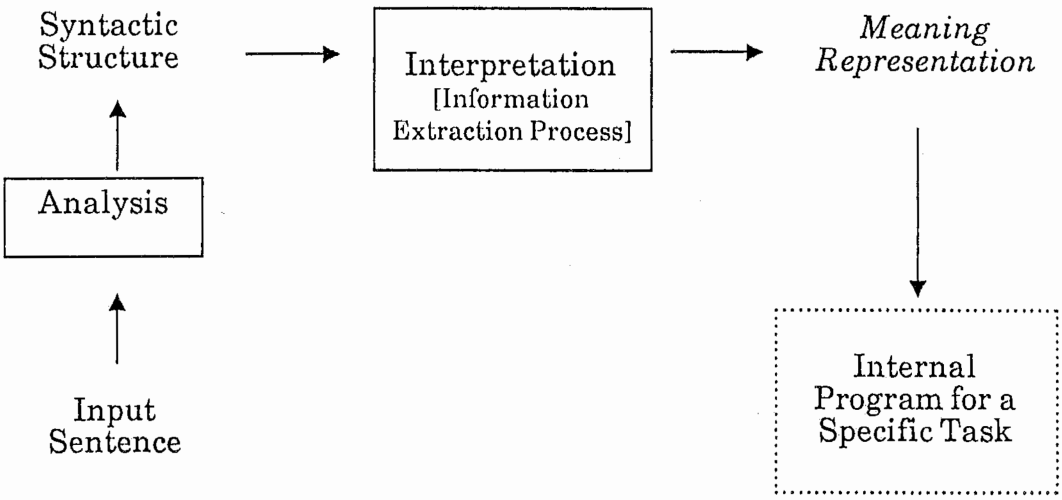


Fig. 1 Simplified Framework of NLU

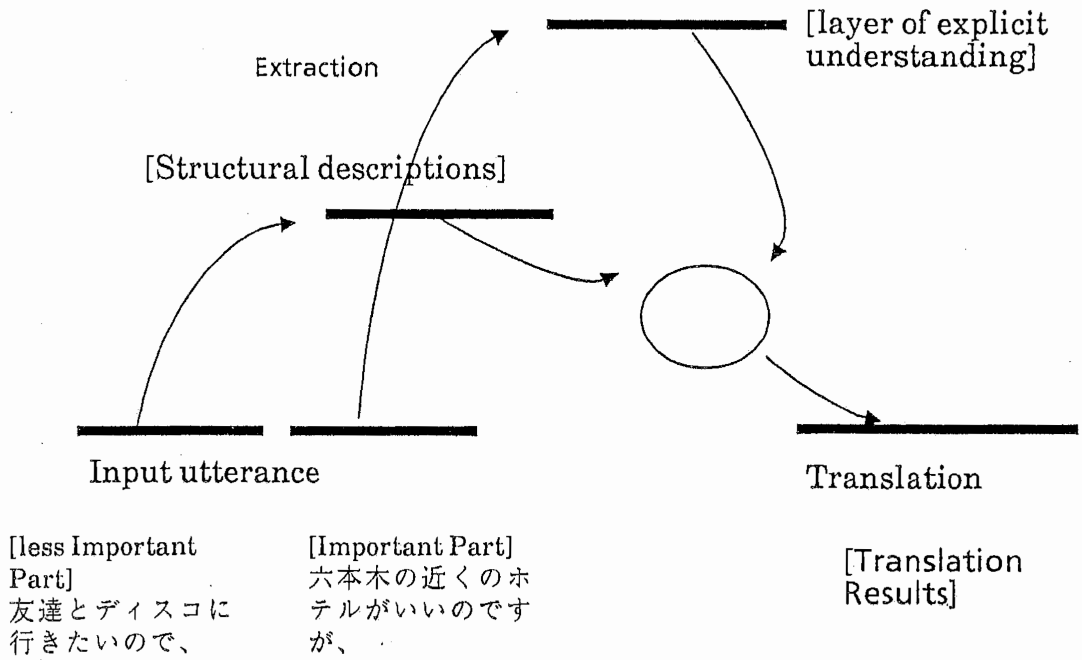


Fig. 2 Schematic view of a DT system

(a) treatment of information

J: 友達とディスコに行きたいので、 六本木の近くのホテルがいいのですが、

E:Because I would like to go to a discotheque with friends, I prefer to stay at a hotel near Roppongi.

:less important
: important

(b) varieties of surface expressions

六本木の近くのホテルがいいわ。
[structure bound translation] a hotel near Roppongi is good.
六本木あたりのホテルをお願いします。
[structure bound translation] a hotel around Roppongi, please.
ホテルは六本木の近くがいいのですが、
[structure bound translation] as for hotel, in the vicinity of Roppongi is good.
都合がいいのは六本木に近いホテルです。
[structure bound translation] what is convenient is a hotel near to Roppongi.

the same
meaning
representation

(c) fluent translation only for important part

[Friends Discotheque 'to go'] I prefer to stay at a hotel near Roppongi.

Fig.3 Examples of Dialogue Translation Processing

AN APPLICATION OF PRAGMATIC KNOWLEDGE TO A BOTTOM-UP ANALYSIS OF JAPANESE SENTENCES

Tadahiro Kitahashi and Makoto Hirai
The Institute of Scientific and Industrial Research
Osaka University

1. Introduction

A sentence or utterance in a discourse can not be fully understood in isolation since it is both syntactically and semantically incomplete because of such linguistic phenomena as ellipsis and anaphora. Resolution of ellipsis and anaphora can be thus the first step to understanding of texts or dialogue. Another difficulty in Japanese analyses arise from the vague boundary between a sentence and texts. A Japanese sentence can be a paragraph in effect. Determining the dependency relations among the simple sentence in a sentence is almost as difficult as understanding texts. This paper discusses an application of pragmatic constraints which underlie utterance and connective expressions.

2. Ellipsis resolution

In Japanese, obligatory case elements and noun phrases modifying function nouns, which will be explained later, are often omitted. When considering the construction of an intelligent man-machine interface via natural language, it is of great importance to determine what words are omitted. One approach to this problem is hypothesis-driven analysis, that is, using knowledge about individual situation, such as "script". This approach can be efficient insofar as the theme or the situation of a discourse can easily be reminded of. However, it requires relatively large amount of knowledge about conventional or causal relations among individual events.

We outline in this section, through an example, another approach to ellipsis resolution when communicative or cognitive verbs, such as KIKU (ask), OSHIERU (tell) and SHIRU (know), are used in utterance. It is based upon pragmatic constraints imposed on omitted words. The pragmatic constraints are derived from the lexical information together with pragmatic assumptions underlying utterance. The approach allows to correctly supplement such omissions as in the example below. They can not be supplemented syntactically or semantically. Note that the words in brackets do not appear in Japanese.

ex1) TARO-HA HANAKO-NI JUSHO-WO KII-TA-GA
OSHIE-TE KURE-NAKATTA.
(Taro asked Hanako her address.
But [She] did not tell [it to him])

2.1 Lexical Information

First we explain the lexical information of verbs and nouns necessary for the discussion here. Two semantic categories of noun, "function" and "event", are of primary importance. Nouns representing attributes, such as "height" and "color", and inter-event relations like "reason" and "purpose" are classified as functions with one argument and one value. For example, function "height" takes a physical object as the argument and gives a number as its value. Function "reason" takes a fact as its argument and gives another fact as its value as well. Nouns representing action, phenomenon or state are classified as events. Simple sentences comprising a sentence also belong to events. Events are further divided into "facts" and "propositions" according to their truth values.

As for verbs, four semantic attributes, controllability, structure of transfer (SOT), operativity and semantic constraints on case elements are useful for supplementing ellipsis. Controllability specifies that a verb represent a volitional action or not. SOT is a structural description of verb representing transfer, appearance and disappearance of some object and is represented in a list of four elements, each of which is a case comprising the case frame. Following three examples are the typical examples of SOT.

1. X-GA Y-WO Z-NI KIKU (X asks Y to Z). SOT of KIKU (ask) is (WO GA NI -),i.e. transfer of Y from X to Z.
2. X-GA Y-WO SHIRU (X know Y). SOT of SHIRU(know) is (null - WO GA),i.e. appearance of Y at X.
3. X-GA Y-WO WASURERU (X forget Y). SOT of WASURERU (forget) is (null WO - GA), i.e. disappearance of Y from X.

Operativity (OP) represents affecting object and affected object in the case elements of a verb as described below.

4. X-GA Y-WO NAGURU (X beats Y). OP of NAGURU(beat) is (GA WO), i.e. X directly affects Y.

In a communicative verb, semantic constraints on the (transferring) object is usually either a function value or an event. Especially for a verb like "ask", no fact is allowed as its object. In a cognitive verb, the object is always a fact or a function value. Other case elements in both types of verb are normally persons.

2.2 Supplementing Omitted Words

In order to supplement correct words in the example, the knowledge of "who knows what" and "who does not know what" is necessary. In other words, pragmatic rules in utterance are to be taken into account.

"Asking" means that an agent requires the transfer of some information from a recipient. In such a situation, we can assume that the agent does not know the information while the recipient knows it. "Telling" is the transfer of some fact from an agent to a recipient. Here, we can also assume that the agent knows the information while the recipient does not know it. We can further assume that any person knows his/her own function values, such as his/her address. From these assumptions, we can derive Taro does not know Hanako's address and Hanako knows it. Since "address" is a function in the example, we can also derive Taro and HANAKO know their own addresses. These are what we call the pragmatic constraints. By assigning semantically adequate candidate to each of the omissions so that no contradiction arise among the assumptions and the sentence, we reach the conclusion that the second simple sentence says "Hanako did not tell her address to Taro." Thus the omitted words are identified.

2.3 Derivation of Pragmatic constraints

It is plausible that the assumptions listed below underlie every utterance insofar as the situation is normal. These assumptions are obviously independent of specific domain and task.

1. Nobody can tell another what he/she does not know.
2. Everybody knows his/her own function value.
3. The agent of an action knows the action.
4. The person involved in an event knows the event.
5. A recipient of a object knows the object.

By combining these assumption with lexical information of the verb in a sentence, pragmatic constraints like those used in the example can easily be derived. As an example, if a verb has (WO GA NI -) as its SOT and the case element for WO-case is fact(s) or function, the verb represents transfer of fact. Consequently, assumption 1 and 5 are applicable. This gives us the pragmatic constraints used in the example.

3. Complex Sentence Analysis

3.1 Semantic constraints on sentence structure

A Japanese sentence generally consists of a sequence of simple sentences connected by connective expressions. Some connectives require specific semantic attributes of both superordinate and subordinate sentences. As an example, temporal connective NAGARA, which means simultaneous proceeding of two events, requires both sentence be non-stative events. In ex2), MI(watch)-NAGARA does not depends upon (or modifies) YOKU-NAI (not good) but upon YOMU (read), because "watch" and "read" are both non-stative events while "not good" is stative. In ex3), for the same reason, ARUKI(walk)-NAGARA depends upon SAGASU (look for), since both are non-stative events.

ex2) TEREBI-WO MI-NAGARA HON-WO YOMU SHUUKAN-HA YOKU-NAI.
(It is not a good habit to read a book while watching TV.)
ex3) GAIRO-WO ARUKI-NAGARA ENZETSU-NI TEKISI-TA BASYO-WO SAGASU.
(Walking the street, I look for a good place for a speech.)

Semantic constraints imposed on sentence constituents by a connective can, thus, often determine the structural difference between syntactically equivalent sentences as explained above.

3.2 Pragmatic constraints underlying connectives

Such constraints are, however, too weak to determine the relations among the constituents of a sentence. Here, we can observe another pragmatic constraint which is concerned with the relationship between the use of a connective and the assumptions the speaker has in mind. When a speaker uses a connective which represents "reason" or "cause", such as NODE, TAME, he/she has assumed that two events (sentences) connected by the connective are "facts" even if they will occur in the future. On the other hand, when he/she uses a connective representing condition, such as TO(when) or TEMO(even if), two events connected by the connective are assumed to be "propositions", not specific "facts". Most of the other connectives, for instance NAGARA mentioned above, do not reflect such speaker's assumptions. Whether a event (sentence) with which this kind of connective is accompanied is "facts" or "propositions" is totally dependent on whether superordinate sentence is a "fact" or not.

ex4) KOUSHA-WO SINCHIKUSURU-TAME GAKUSEI-GA HUE-TEMO SHINPAINAI.
(We'll build a new school building. So, you don't have to worry even if the students grow in number.)

Dependency relations among simple sentences in typical complex sentences can be determined by these constraints along with semantic ones. For example, the structure of the sentence ex4), which is both syntactically and semantically ambiguous, can be uniquely determined as shown in Fig.1. Note that the partial sentence E_1 in the figure is regarded as a fact.

```
(build a new school house)   (grow in number)   (do not worry)
KOUSHA-WO SHINCHIKU-SURU     GAKUSEI-GA HUERU     SHINPAI-NAI
          |fact                    |prop              |prop
          |                        |__________________|
          |                                 |TEMO(condition)
          |                        E1(fact)|
          |________________________________|
                         |
                        fact
```

Fig.1 Dependency relations of example 4

4. Concluding Remarks

We have outlined the pragmatic constraints underlying utterance and use of connectives. As can be seen clearly, these constraints do not depend upon a specific domain or task. Hence, it provides a uniform mechanism for ellipsis resolution and complex sentence analyses.

References

1)S.Kuno : "Nihon Bunpou Kenkyuu (Studies of Japanese Grammar)", Taishukan, 1973, Tokyo.
2)M.Brady and R.C.Berwick : "Computational Model of Discourse", MIT press, 1983.
3)J.Hinds and I.Howard : "Problems in Japanese Syntax and Semantics", Kaitaku-sha, 1979, Tokyo.
4)M.Hirai and T.Kitahashi : Determination of Omitted Words in Japanese Sentence by Pragmatic Rules, Proc. of the 1st Ann. Conf. of JSAI, 1987, pp377-380 (in Japanese).
5)M.hirai and T.kitahashi : "Analysis of Japanese Complex Sentences Using Factuality of Events", Proceedings of the conference of Natural Language Processing Interest Group of Japan Information Processing Society, Sept, 1987.

Speech Act Theory and Extended Discourse

C. Raymond Perrault
Artificial Intelligence Center and
Center for the Study of Language and Information
SRI International

Introduction

In its early days, the central task of speech act theory was generally conceived as that of explaining what actions a speaker was performing in using language. Recent work by Cohen and Levesque [Cohen and Levesque 1987a, b] and the author [Perrault 1987] has focused instead on relating utterances to changes in the mental state of the participants. We will review this work and discuss some of its implications for semantics and pragmatics.

Austin's Categorization of Speech Acts

Austin [Austin 1962] proposed three levels of description of speech acts. In saying "It's cold here," for example, the speaker utters a well-formed sentence of English with definite sense and reference. This act *of* saying something Austin dubs the *locutionary* act. In so saying, he may also be asserting that it is cold, or even indirectly requesting the hearer to turn the heat up. These actions performed *in* saying something Austin calls *illocutionary* acts. Finally, actions performed *by* saying something are called *perlocutionary* acts. One important feature of illocutionary verbs is that they can be used in so-called explicit performative sentences such as "I hereby *assert* that it is cold here," whereas perlocutionary verbs cannot (e.g. * "I hereby *convince* you that it is cold here").

Two Approaches to Speech Act Theory

There can be two kinds of approaches to speech act theory. One is the *primacy of action approach* which was taken, for example, by Searle [Searle 1969]. In this approach, speech act theory seems at first glance to reduce to the problem of characterizing the successful performance of illocutionary acts, i.e., of giving conditions under which the performance of action, and in particular the utterance of one or more sentences, can be taken to be the successful performance of a given speech act. However, this approach has limits in explicating the relation between an utterance's form and its function. As illocutionary acts can be performed in various ways, i.e., *mistakenly* (as in assertions whose content happens to be false), *insincerely* (as with lies), *indirectly* (as in using "It's cold here" as a request to turn up the heat) and *non-seriously* (as in "This is the best meal I've ever had" to mean that it is terrible), it is difficult for this approach to give direct definitions relating the form of utterances to the illocutionary acts they are used to perform.

The other approach is the *primacy of mental state change approach*, which is taken in recent work by Cohen and Levesque [Cohen and Levesque 1987a, b] and the author [Perrault 1987]. The following characterizations are given in this approach:

- chracterization of mental state change associated with utterance in virtue of its features (+ context),
- characterization of (non-institutional) illocutionary acts independently of utterance features,
- characterization of perlocutionary acts in terms of change in mental state.

In [Perrault 1987], many of the consequences of speech acts are defined not only *axiomatically* but as *defaults* which are assumed to hold as long as there is no evidence to the contrary. This default theory describes the following:

- the change in attitude of an agent over time,
- the transfer of attitudes between agents,
- the consequences of an action for the mental state of an observer,
- the assumption that actions are intentional,
- the assumption that the utterance of sentences with particular features reveals particular aspects of the mental state of the speaker.

The restriction of the realm of semantics to truth-conditional aspects of meaning in traditional semantic theories has precluded a uniform semantic treatment of declarative, imperative and interrogative sentences, and has left out explicit performative sentences. However, this default theory offers uniform treatment of mood, compatible with relational theory of meaning at the same time allowing interpretation of an utterance to be affected by participant's mental state (e.g. "Where is the bank?" is interpreted differently depending on the recipient's mental state).

Axioms and Default Rules for Speech Acts

Note how a set of default rules for reasoning in speech act theory can be stipulated on the basis of Reiter's Default Logic [Reiter 1980] (limiting ourselves here to declarative sentences and also disregarding participant's intention).

First, a language L to express facts about agents, actions and propositional attitudes is defined. L has expressions of type *Agent*, *Time*, *Action* and (*Untensed* or *Tensed*) *Proposition*. The constants S and H and the variables x and y are the only expressions of type *Agent*. The integer constants 0, 1, ... and the variables t and u are of type *Time*. If p is of type *Proposition*, then p. is of type *Action*, and denotes the action of uttering a declarative sentence with propositional content p. If x is of type *Agent*, Obs(x) is of type *Action*, denoting the action of observing x. The expressions of type *Untensed Proposition* consist of propositional variables r, s, ... and their boolean combinations. The expressions r_t, $DO_{x,t}\ \alpha$ and $B_{x,t}\ p$ are of type *Proposition*, where r is of type *Untensed Proposition*, p is of type *Proposition*, t is of type *Time*, x is of type *Agent* and α is of type *Action*. They are read as "r is true at time t," "x did α at time t," and "x believes at time t that p" respectively.

Beliefs are constrained by both axioms and default rules. We assume that an agent's beliefs at any one time follow the standard weak S5 *axioms*:

[Consistency]	$\vdash B_{x,t}\,p \supset \neg B_{x,t}\neg\, p$
[Closure]	$\vdash B_{x,t}\,p \,\&\, B_{x,t}(p \supset q) \supset B_{x,t}q$
[Positive introspection]	$\vdash B_{x,t}\,p \supset B_{x,t}B_{x,t}\,p$
[Negative introspection]	$\vdash \neg B_{x,t}\,p \supset B_{x,t}\neg B_{x,t}\,p$
[Necessitation]	If $\vdash p$ then $\vdash B_{x,t}\,p$

Over time, we assume agents remember their previous beliefs and continue to hold them.

[Memory] $\vdash B_{x,t}\,p \supset B_{x,t+1}B_{x,t}\,p$

The important thing assumed here is the persistence of belief.

[Persistence] $\vdash B_{x,t+1}B_{x,t}\,p \supset B_{x,t+1}\,p$

The belief that an action has been performed is acquired from the observation of an action. The observability axiom is oversimplified as we take agents to be observed, rather than actions. Observing one agent is assumed to imply observation of all actions performed by him.

[Observability] $\vdash DO_{x,t}\,\alpha \,\&\, DO_{y,t}Obs(x) \supset B_{y,t+1}DO_{x,t}\,\alpha$

Two *default rules* are necessary. They both address how new beliefs can be added to old ones. The first allows for transfer of beliefs from one agent to another, as long as the new beliefs are consistent with the old ones.

[Belief transfer rule] $$\frac{B_{x,t}B_{y,t}\,p : M\,B_{x,t}\,p}{B_{x,t}\,p}$$

the rule is read: $B_{x,t}\,p$ can be assumed if $B_{x,t}B_{y,t}\,p$ is believed and $B_{x,t}\,p$ is consistent with what is believed.

The second default associates the utterance of sentences having particular linguistic features with an aspect of the mental state of the speaker — for declarative sentences, that he believes the propositional content.

[Declarative rule] $$\frac{DO_{x,t}\,(p.) : M\,B_{x,t}\,p}{B_{x,t}\,p}$$

One more piece of machinery needed is the following meta-rule which mirrors the *closure* of the axiom system under beliefs:

For all agents x and time t, if $\frac{p : M\,q}{q}$ is a default rule, so is $\frac{B_{x,t}\,p : M\,B_{x,t}\,q}{B_{x,t}\,q}$

These axioms, default rules and the meta-rule can account for various defective, non-serious, and indirect uses of declarative sentences as well as sincere assertions. Because the relation between utterance and the attitude it expresses is given as a default, and thus is defeasible, there is no need to give anything more complicated.

Conclusion

It has been argued that uniform account of speech acts can be given based on the primacy of mental state change approach which regards the 'meaning" of an utterance as the fuction from mental state to mental state. In this approach, speech act theory is only a part of semantics, though it is usually considered as a part of pragmatics. This approach is believed to be the only way to a systematic theory of extended discourses regarded as coordinated rational actions of participants.

References

[Austin 1962] Austin, J.A., *How to Do Things with Words*. New York: Oxford University Press, 1962.

[Searle 1969] Searle, J.R., *Speech Acts*. New York: Cambridge University Press, 1969.

[Cohen and Levesque 1987a] Cohen, P.R. and H. Levesque, "Persistence, intention and commitment," in [Cohen, Morgan and Pollack (eds.) 1987].

[Cohen and Levesque 1987b] Cohen, P.R. and H. Levesque, "Rational interaction as the bases for communication," in [Cohen, Morgan and Pollack (eds.) 1987].

[Perrault 1987] Perrault, C.R., "An application of Default Logic to Speech Act Theory," in [Cohen, Morgan and Pollack (eds.) 1987].

[Cohen, Morgan and Pollack (eds.) 1987] Cohen, P., J. Morgan and M. Pollack (eds.), *SDF Benchmark Series: Plans and Intentions in Communication and Discourse*. Cambridge, MA: MIT Press, 1987.

[Reiter 1980] Reter, R., "A logic for default reasoning," in *Artificial Intelligence*, 13, pp. 81-132, 1980.

(*Summarized by H. Maeda*)

Knowledge-Based Machine Translation

Jaime G. Carbonell and Masaru Tomita

Center for Machine Translation, Carnegie Mellon University

[1] The CMT (Center for Machine Translation)

Of the various engineering approaches to machine translation, the CMT focuses on a knowledge-based approach with the following capabilities:

(a) multi-lingual translations, and

(b) factoring domain semantics and linguistic information.

Along with this approach, the CMT's long-term, practical directions are:

(a) treating real-time dialogue (eg. by telex)

(b) integrating MT with speech processing

Their technological spinoffs expected are:

(a) multi-lingual data-base access

(b) text scanning and summarization

[2] The Universal Parser

The universal parser has been designed to build an MT system with the following capabilities:

(a) multi-lingual, multi-domain translation

(b) domain-independent linguistic grammar for each language, based on LFG/UG

(c) language-independent semantic domain knowledge, based on hierarchical frame methods

(d) easily extensible to new language and domains

(e) sharing a common source grammar between generation and parsing

A machine translation system based on the universal parser adopts the pivot method and, then consists mainly of a run-time parser and generator. The pivot method requires only *2N* parsing and generation grammars for MT of *N* languages while the transfer method requires *N(N-1)* transfer grammars. The run-time parser takes as its input a sentence and returns a domain-dependent semantic representation. The generator takes a semantic representation and returns a sentence. For MT efficiency, source grammar is compiled into efficient semantic/syntactic grammars for parsing and generation. These are used by the run-time parser and generator, respectively.

Grammar Compiler: The universal parser uses separate syntactic and semantic knowledge and lexicons in order to apply multi-lingual and multi-domain MT easily. These are compiled into an efficient semantic/syntactic grammar by the grammar compiler through the following processes:

(a) Syntactic knowledge in LFG-like notation, lexicons and domain semantic specifications in a frame representation language are mixed and compiled into syntax/semantics grammar in LFG notation.

(b) The LFG grammar is compiled into an augmented CFG.
(c) The augmented CFG is compiled into an augmented LR parsing table. The table is used by the run-time parser.
(d) Syntax/semantics grammar in LFG are also compiled into an augmented CFG for generation, i.e., in the universal parser, both analysis and generation use the same source grammar.

Run-Time Parser: The run-time parser uses an augmented LR parsing table to translate input sentences into their semantic representations. The parser uses the generalized LR parsing algorithm for augmented CFGs.

By using the universal parser, the following have been implemented in Common Lisp:

(a) English to Japanese,
(b) Japanese to English,
(c) Japanese to German, and
(d) English to German.

[3] Parsing Noisy Sentences

The CMT developed a method to parse a noisy sentence, one which might include errors due to a speech recognition device. The parser based on this method is connected to a speech recognition device which takes a continuously spoken sentence in Japanese and produces a sequence of phonemes. The generalized LR parsing algorithm is adopted and modified to handle errors, that is, altered phonemes and extra and missing phonemes, as follows:

To treat altered phonemes: The parser dynamically create a phoneme lattice by placing alternate phoneme candidates in the same location as the original phonemes. Each possibility is explored by each branch of the parser. Information on alternate phoneme candidates is obtained from a confusion matrix.

To treat extra phonemes: The parser has one branch for each extra phoneme candidate and ignores the phoneme. The parser assumes that, at most, only one extra phoneme can exist between two real phonemes.

To treat missing phonemes: The parser inserts possible missing phonemes between two real phonemes. The parser assumes that between two real phonemes, at most, only one phoneme can be missing.

The parser employs a scoring method using the confusion matrix:

(a) to prune the search space by discarding branches of the parse with hopelessly low scores.
(b) to select the best sentence from multiple candidates by comparing their scores.

The parser has been implemented in Common Lisp and is being integrated into CMU's knowledge-based machine translation system to accept a spoken Japanese sentence in the domain of doctor-patient conversation and generate sentences in English, German and Japanese.

(Summarized by K. Kogure)

THE SYSTEM OF INTELLECT AND EMOTION : AN ENGINEERING APPROACH TO THE MODEL OF MIND

Naoyuki OKADA

Department of Information Science
and Systems Engineering
Oita University
Oita 870-11, JAPAN

M. Minsky recently proposed a model of mind, "The Society of Mind." This paper makes another proposal for the model of mind and has two features. One is engineering approach : We aim to make a computer system for comprehensive and unified understanding of natural language, picture patterns and voice. The other is introduction of emotional functions to the system : We regard intellect and emotion as the two wheels of a cart in mind.

1. Hierarchy of the system of intellect and emotion

Mind consists of five levels along the abstraction process of knowledge.

Level 1 Raw data

Raw data is sensed by the five senses and is nearly a copy of objects (things, events and attributes) in the real world. We consider some of mental objects are also sensed by something like internal sensors. Raw data has a coded structure like a digitized image.

Level 2 Cognitive features

They are features extracted from raw data. The so-called feature extraction is done at this level. The feature extraction of a change from one state to another is important in events, while that of a difference between two objects is important in attributes.

Level 3 Conceptual features

They are divided into two classes. Class 1 is symbolic data associated with cognitive features. Some of semantic markers assigned to words are regarded as being of this type. Class 2 is symbolic data which function as connectors of concepts at level 4 or 5.

Level 4 Simple concepts

These concepts are composed of conceptual features. Some of them are primitive and are not reduced into further elementary concepts. Many of

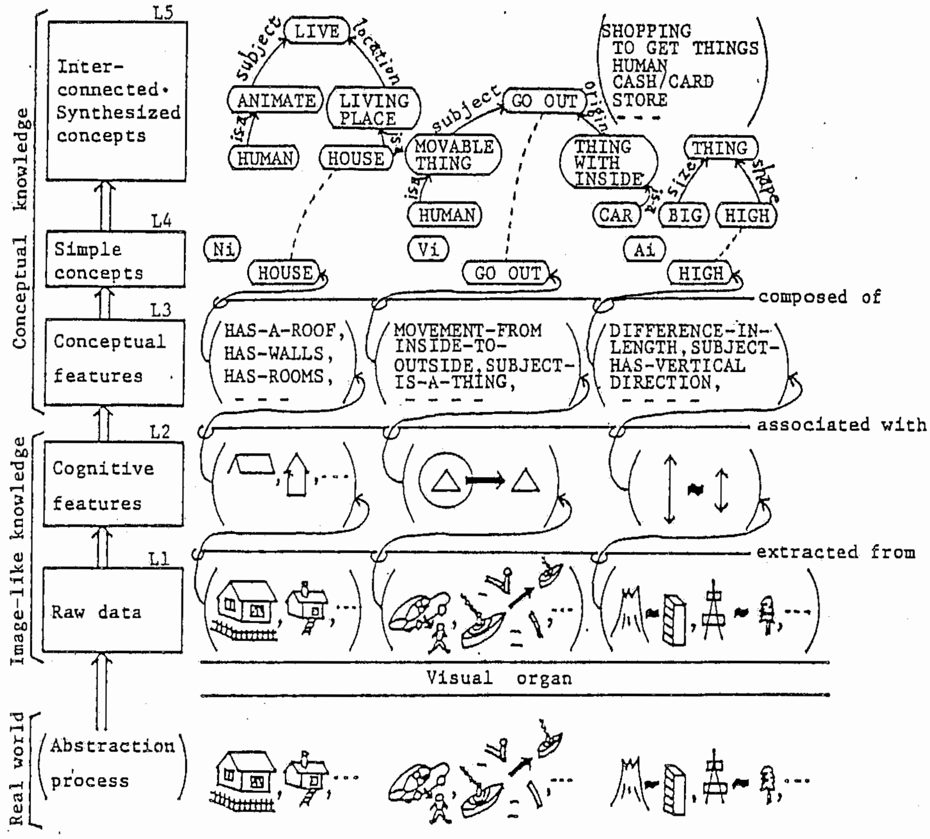
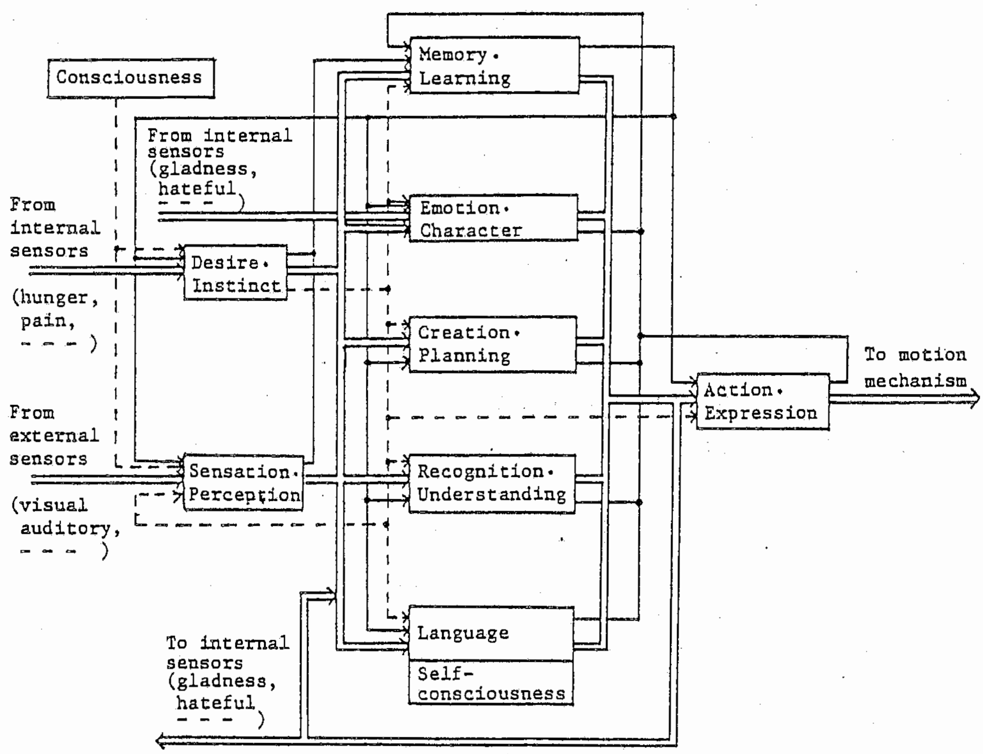


Fig.1 Hierarchy of the system of intellect and emotion



⟹: Flow of process, ⟶: Flow of data, --→: Flow of control

Fig.2 Domains of the system of intellect and emotion them have names as words. In the case of verbs and adjectives, they have a "case frame."

Level 5 Interconnected * Synthesized concepts

These concepts are obtained by interconnecting or synthesizing simple concepts. "Semantic network" and "scripts" are regarded as examples of this kind of interconnection. Some of them have names as words.

A rough sketch of this hierarchy is shown in FIGURE 1. In the hierarchy mentioned above, mental data has two types, declarative and procedural.

2. Domains of the system of intellect and emotion

Mind comprises eight domains according to mental activities.

Domain 1 Sensation * Perception

Domain accepting external data by the five senses at level 1 and extracting cognitive features from accepted data at level 2. Parallel processings are done in this domain.

Domain 2 Recognition * Understanding

Domain making a given data correspond to concept data at level 4 or 5.

Domain 3 Desire * Instinct

Domain accepting internal data such as hunger, fatigue or sexual feeling by internal sensors at level 1 and arousing desires such as meal, rest or sex respectively, at level 2. Let's call the aroused desire "intention." It has two functions : one is to define goals to satisfy the desire and the other is to define criteria to select a desirable action among alternatives.

Domain 4 Creation * Planning

Domain arranging physical and mental activities to carry out a given goal at levels 3 through 5. The so-called problem solving is done in this domain.

Domain 5 Emotion * Character

Domain evaluating the value of a given data by internal sensors (measures) such as gladness or hatefulness at levels 3 through 5. Emotion is regarded as evaluation of data with measures, and character as pattern of evaluations among emotional measures.

Domain 6 Action * Expression

Domain making preperations for physical movements of body according to the results of creation * planning at levels 5 through 1.

Domain 7 Memory * Learning

Domain supplying necessary data to the processing in the other domains and accepting the processed data in the other domains.

Domain 8 Language

Domain analyzing and generating sentences at levels 1 through 5. The essential function of language is to describe every data or processing in every domain. From this derive the following three functions:

(1) To communicate with others,

(2) To get one's ideas or inferences into shape,

(3) To observe every data or processing in every domain.

In particular, the function (3) is closely related with "selfconsciousness." The observation lets one know one's own mental activity.

A rough sketch of the domains is shown in FIGURE 2. Domains may overlapp one another although they are drawn apart in FIGURE 2.

3. Unified understanding of natral language, picture patterns and voice

Let's first consider picture pattern underdstanding and natural language expression.

Given an input picture patern, each object in it is first associated with data at level 1 which are stored in the domain of *memory * learning.* The objects may be things, events and attributes. The raw data obtained are then associated with data at level 2 by feature extraction. The processing so far is performed in the domain of *sensation * perception.*

Features extracted are further associated with data at level 3, and thus objects in the input are eventually associated with interconnected * synthesized concepts at level 5, which are also stored in the domain of *memory * learning.*

These manipulations at levels 3 through 5 are performed in the domain of *recognition * understanding.* Here assosiation of the data at one level with that of another is called "interpretation."

Next, the data interpreted at level 4 or 5 are mapped to linguistic data at level 4 or 5 in the domain of *language* by semantic processing.

The linguistic data mapped have the so-called deep structures, which are then transformed to surface structures by syntactic processing. Thus, sentences representing the input are obtained in the domain of *language.*

The sentences are processed in *action * expression* to utter or write down them. As a result, declarative and procedural data are arranged at level 5 through 1. The declarative data at level 1 may be the contents uttered or written down and the procedual data at level 1 may control the motions of one's mouth or hand. Finally, sentences are uttered or written down. Interpretation thus proceeds also in the downward direction.

A rough sketch of the whole process is shown in FIGURE 3.

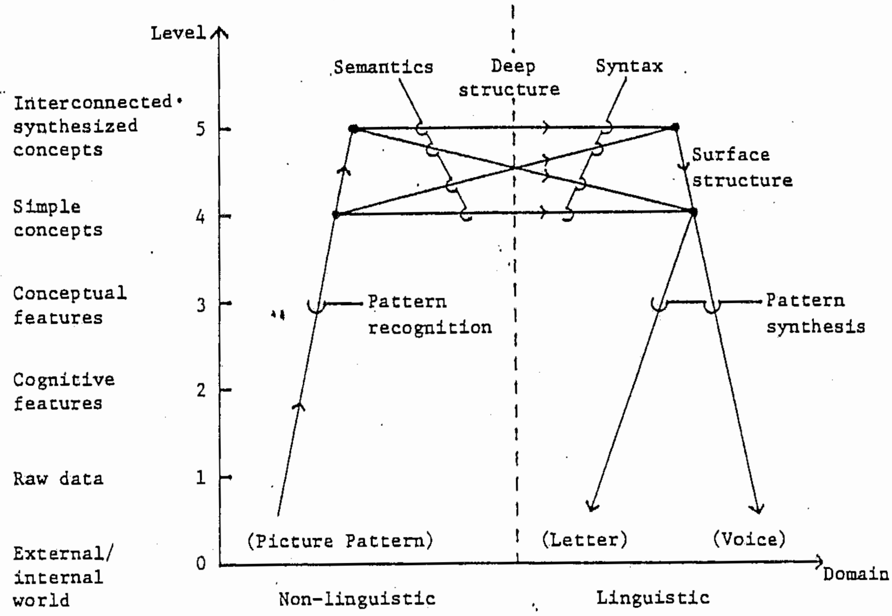


Fig.3 Process of understanding and expression

On the other hand, when the input is sentences, interpretation proceeds in the reverse direction of FIGURE 3. For example, a sentence written down is first processed as character recognition at levels 1 through 3. Then the linguistic data with surface structures at level 4 or 5 are obtained. Next, the linguistic data are mapped to non-linguistec data in other domains through deep structures. Sometimes interpretation proceeds downward to level 1. The concrete representations involved in this process will be accompanied by visual images.

In interpretation mentioned above, mind will face ample ambiguity since there are several levels and several domains, and furthermore there are a lot of chunks of knowledge in each domain. Some ambuguities may be dissolveded by natural constraints among data, and some by introducing heuristics. There will still remain ambiguities in the procedure.

Let's call an interpretation "suitable" if it suits an intention. This leads to the following definitions : an "understanding" is a suitable interpretation in the upward direction and an "expression" is a suitable interpretation in the downward direction.

As shown in FIGURE 3, both understanding and generation of natural language, picture patterns and voice can be manipulated in a comprehensive, unified way by the system of intellect and emotion proposed.

THESAURUS CONSTRUCTION FROM A JAPANESE LANGUAGE DICTIONARY

Hiroaki TSURUMARU (Nagasaki Univ.)

1. INTRODUCTION

How to obtain hierarchical relations(e.g.superordinate-hyponym relation(>),synonym relation(≡))is one of the most important problems for thesaurus construction. We have been developing a system for extracting these relations semi-automatically from a machine readable dictionary (MRD : Shinmeikai Kokugojiten, published by Sansei-do, in machine readable form).

In this paper,the several features of definition sentences{DS} and the mechanical extraction of the (hierarchical) relations between words are discussed.

2. FEATURES OF DS IN ORDINARY JAPANESE LANGUAGE DICTIONARY

2.1 STRUCTURE OF DS

The features of DSs in the dictionary are as follows:

(a) Generally, DS of an entry word{EW} is defined by qualifying its superordinate word or synonyms or hyponyms. We call these words definition words{DW}.

(b) Generaly, the final word in DS is DW.

(c) In some cases,the final expression in DS assignes semantic relation between DW and EW. We call this expression functional expression{FE}. DW is just before the FE.

(d) Genraly, DW is modified by another phrase(modifier).

(e) In some cases, DS contains more than one DW.

According to these features the following general structure is obtained .

([MODIFIER] + DW) # + [FE] 。

Notes) [⋯] : optional constituent

(⋯) : required constituent

\# : sequence of coordinate constituent(e.g. ・, と, や)

\+ : concatination symbol

2.2 FEATURES OF FE

FEs prescribe hierarchical relations (DW>EW, DW<EW, or DW≡EW)

or the whole-part relation (DW》EW). Besides these rela-tions, another relation between DW and EW is prescribed by certain FEs called associative relation(R).

There are so many FEs that they are mainly divided into four patterns called functional patterns{FP: a brief notation}. FP is expressed by means of regular expression. FP is necessary for extracting FE and DW-EW relation information (i.e. information necessary for deciding the DW-EW relations) assigned by FE. FP also designates a place of DW in DS.

Main features of FP are as follows:

(1) Type100 : 「(…DW)#」 + σ * + FW

(2) Type200 : (…DW)# + の + FW

(3) Type300 : (…DW)# + P + FW

(4) Type400 : (…DW)# + など

Notes) σ * is arbitrary character string,
P is special phrase(e.g. に対する),
\+ is concatination symbol.

We collected about one hundred seventy FWs. These are classified into two groups. In one group (contained 64 FWs), the FWs contain explicit DW-EW relation information. In the other group (contained 105 FWs), some of the FWs contain usages of the EWs, which are also important to thesaurus.

We have constructed a FW dictionary which includes FP and DW-EW relation information corresponding to FP.

3. EXTRACTION OF DW-EW RELATIONS

3.1 ASSUMPTIONS FOR DECIDING DW-EW RELATIONS

The assumptions are as follows:

① When DS dosn't include FE, DS ≡ EW.

② When DS includes FE, SS ρ_{FE} EW.
SS is standardized DS, that is, ([MODIFIER] + DW) #

③ [MODIFIER] + DW ≦ DW

④ $([MODIFIER_i] + DW_i)^\# \geqq [MODIFIER_j] + DW_j$
Where $i, j = 1 \sim n$.

3.2 ALGORITHM FOR DECIDING DW-EW RELATIONS

The algorithm for deciding the DW-EW relations is obtained by means of these assumptions[2].(see OHP)

4. SYSTEM FOR EXTRACTING (HIERARCHICAL) RELATION

The system consists of the following four steps.

(1) Extraction of EW and its DS from MRD.

(2) Extraction of FE and DW-EW relation information(ρ_{FE})

The FW dictionary is used for extracting ρ_{FE}.

When DS includes FW but doesn't conform FP or when DS includes more than one FW, DS is separated out as check data.

(3) Extraction of DW and DW-EW relation information

A general word dictionary[2] is used for extracting DW.

The relation information is also extraced, that is, 'DW is modified ' and 'The number of DW is only one' .

When DW isn't extracted,the DS is separated out as check data.

(4) Decision of DW-EW relation

According to the above-mentioned algorithm , DW-EW relations are decided.

When extracted relation information is ambiguous, DS is separated out as check data.

5. CONCLUDING REMARKS

(1) System precision(the ratio of correct output data to output data) is about 95% .

(2) The ratio of chech data to input data is about 14%.

(3) Most of incorrect output data occur in the step of extraction of DWs which are described by 'hiragana' notation, because of limitation of the longest matching method.

(4) There is a shortage of semantic information(e.g.lack of the a dequate DW) in the ordinary language dictionary because of the assumed human usage of the dictionary.

(5) Some topics for further discussion are as follows:

① Extraction of hierarchical structure among entry words.

② Analysis of Modifier of DW in DS and extraction of views.

③ A scheme for the representation of the meaning of a word.

A pilot system has been implemented on FACOM M-360 (Nagasaki University Computer Center) by PL/1.

REFERENCES

(1) H.Tsurumaru,et.al.: Proc.of COLING'86, pp.445-447 (1986.8)

(2) H.Tsurumaru,et.al.: FI Technical Report of IPS,3-1(1986.11)

Intonational Meaning in the Interpretation of Discourse

Julia Hirschberg

AT&T Bell Laboratories

It is widely accepted that intonation plays an important role in conveying information about discourse. However, just what this role is, is not well understood. It has proven difficult to provide a precise mapping between intonational features and discourse information, in part due to the lack of adequate systems for intonational description, in part due to the difficulties of representing discourse information, and in part due to the paucity of reliable studies of intonational meaning. In short, the little that is known of intonational meaning has been difficult to describe. Thus, it has been difficult to envision the use of intonational information in computer systems which generate or interpret natural language.

Research now underway at Bell Laboratories and elsewhere is changing this picture. In this talk, a framework[5] for analyzing English intonation was described, which permits a precise description of intonational features such as pitch range, pitch contour, accent placement, and intonational phrasing -- and a comparison of the meanings conveyed by varying the values of these features systematically. And then a number of research projects currently underway which use this framework to identify the discourse correlates of various types of intonational variation was discussed. These studies include work on the conveying of discourse structure intonationally[3,1,2,6], the role of intonation in reference resolution[4], and the intonational disambiguation of direct and indirect speech acts. How variation of intonational features of pitch range and final lowering can communicate different topic structures in a discourse, how differences in accent type and phrasing can distinguish so-called *cue phrases* (words and phrase such as *now, well,* and *by the way* which can be used to communicate discourse structure), how the location of nuclear stress influences the interpretation of *bound anaphors* in sentences like '*Tom likes his work and so does Bill*', and how choice of pitch contour affects the interpretation of speech acts in sentences like '*Can you log in on this machine*' were shown. Findings presented (some of them preliminary) are based on the analysis of recorded naturally occurring speech and a variety of production and perception experiments. (See Fig.1 and Fig.2)

The overall goal of these studies is the definition of relations between intonational features such as pitch range, pitch contour, and accent placement and discourse features such as topic structure, focus, and information status -- as well as the identification of interpretations for particular instantiations of these features. Results are currently being applied to the generation of speech from an abstract representation of the message to be conveyed, in a system which generates spoken directions for drivers in Cambridge MA. While the current application involves speech generation, the theoretical results should also be applicable to speech understanding.

References

[1] J. Hirschberg and D. Litman. Now let's talk about *now*: identifying cue phrases intonationally. In *Proceedings of the 25th Annual Meeting*, Association for Computational Linguistics. Stanford, 1987.

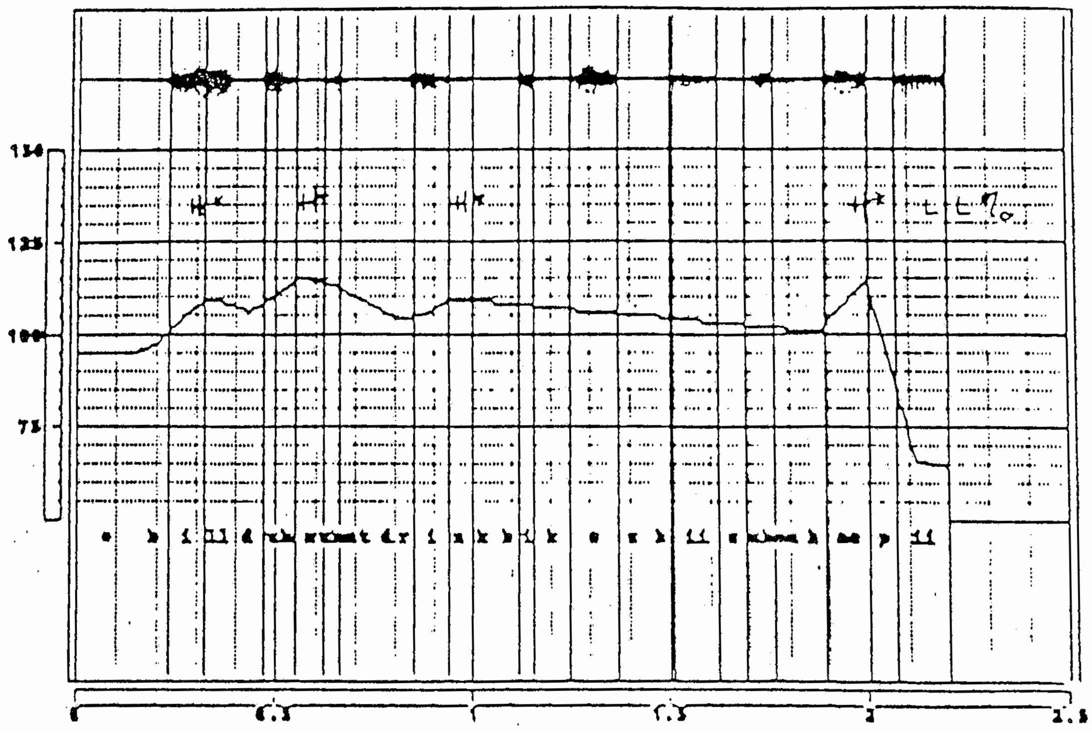
[2] J. Hirschberg, D. Litman, J. Pierrehumbert, and G. Ward. Intonation and the intentional structure of discourse. In *Proceedings*. IJCAI-87, Milan, 1987.

[3] J. Hirschberg and J. Pierrehumbert. The intonational structuring of discourse. In *Proceedings of the 24th Annual Meeting*, pages 136-144, Association for Computational Linguistics. New York, 1986.

[4] J. Hirschberg and G. Ward. Accent and bound anaphora. 1987.

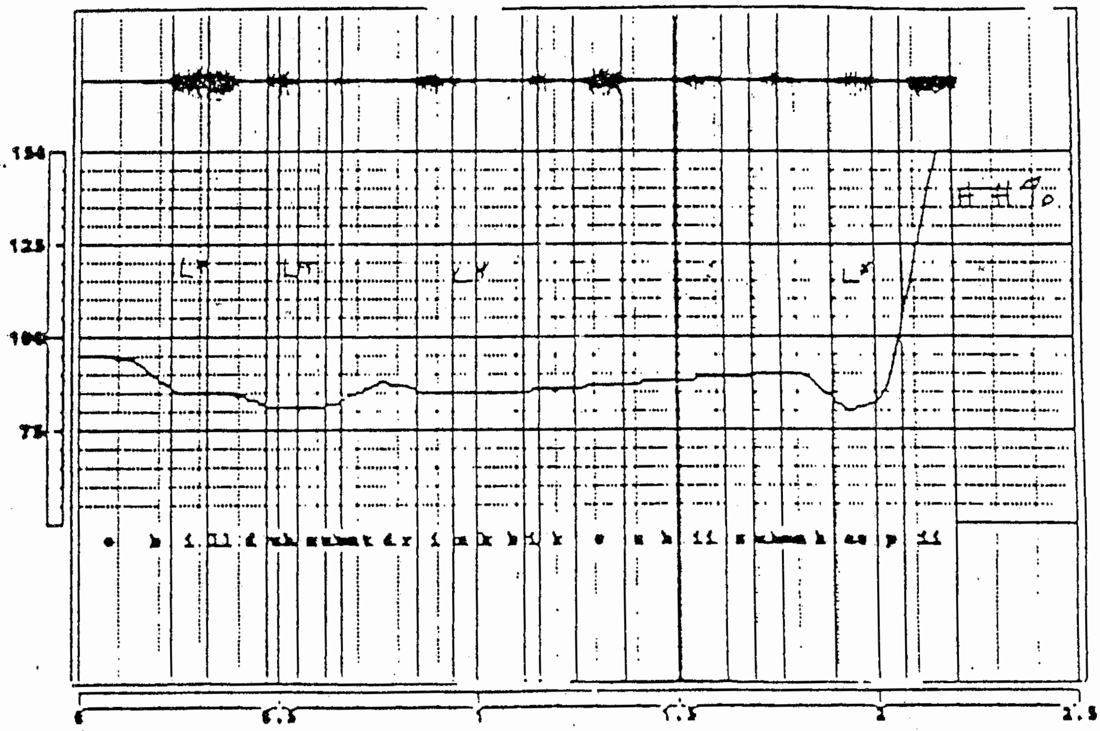
[5] Janet B. Pierrehumber. *The Phonology and phonetics of English intonation.* PhD thesis, Massachusetts Institute of Technology. September 1980.

[6] K. Silverman. Synthesis and perception of paragraph prosody. In *Proceedings.* First Australian Conference of Speech Science and Technology, 1986.



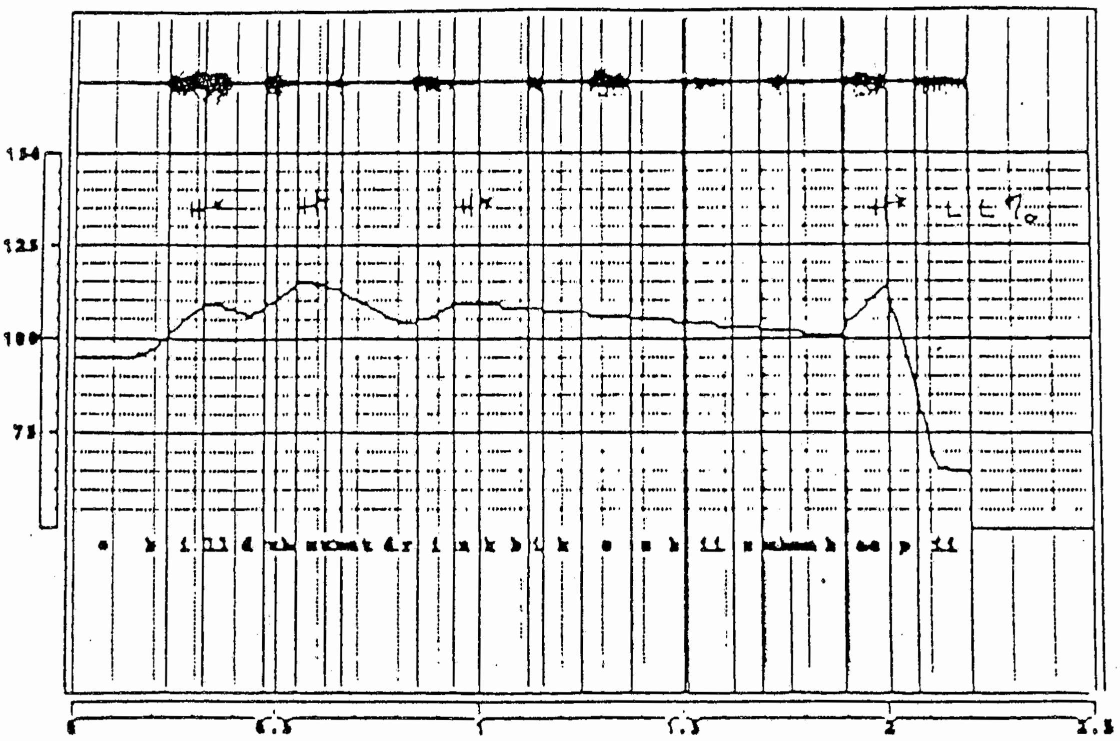
Declarative Contour.

BILL DOESN'T DRINK BECAUSE HE'S UNHAPPY



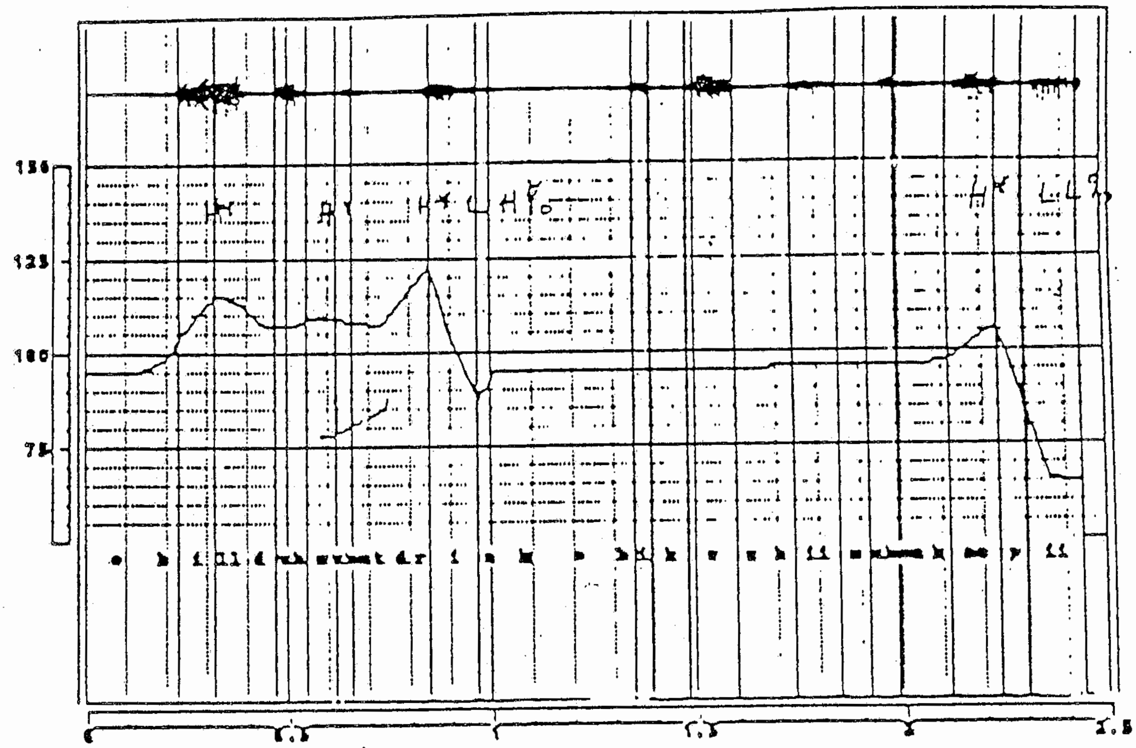
Interrogative Contour

Fig. 1



One Phrase

BILL DOESN'T DRINK BECAUSE HE'S UNHAPPY



Two Phrases

Fig. 2 44

Parsing Natural Language Dialogue

Donald Hindle
AT&T Bell Laboratories

talk at the ATR Workshop on Natural Language Dialogue Interpretation
Osaka, Japan November 28, 1987

This talk presents a theory of natural language parsing, illustrated in terms of a parser for a fragment of Japanese syntax. We assume that the *goal* of parsing is to produce a *description* of an utterance, and the *task* of parsing is to incrementally produce increasingly complete descriptions of an utterance. In producing this description, a number of distinct modules play a role, among them: 1) the module responsible for phrase structure (i. e. responsible for whatever description of phrase structure can be determined given pre-terminal categories and the phrase structure grammar itself), and 2) a module which describes the division of the input utterance into intonational phrases. We will describe a syntactic analyzer for Japanese, explaining the several modules involved and their interaction, concentrating on the function and interpretation of the particles "wa" and "no".

In the course of parsing an utterance, each module can, on the basis of the current description, *add* new predications to those of previous modules, with the constraint that the statements in the description must be added monotonically: no assertions may be withdrawn. Nevertheless, each module is, to some extent, free to interpret the current description in its own terms. Thus, for example, while the phrase structure component may take "dominates" to mean *directly dominates*, later components are free to take that predicate as meaning merely *dominates somewhere*.

The Phrase Structure Component. This model of parsing makes several specific assumptions about the phrase structure component of language. The categorial description of phrase structure is stated in terms of a strict version of X-bar theory: there is a limited set of categories in the phrase structure, including Noun, Verb, Adj, etc. Each of these lexical categories projects a phrase type of a higher bar level up to the maximum. (In this talk we will assume, mostly for ease of exposition, a single bar level in Japanese.)

Information about particular phrase structure categories is declared in two places:

1) for each phrase structure category, there is a single phrase structure rule, specifying what daughters are admissible and what their order is. All daughters must be maximal projections except the head, which is one bar-level lower than the phrase type. There is at most one head of a phrase; but all other constituents may occur any number of times.

2) associated with each phrase structure category there is a (possibly large) set of templates that specify the admissible leading edges of the constituent.

Consider the simplified specification of phrase structure information in (1), which shows three example rules in their basic form along with their expanded form.

$\bar{I} \rightarrow P\ V\ h$	$\bar{I} \rightarrow \bar{P}^{*}\ \bar{V}^{*}\ (I)$
[P +case] [V]	[$\bar{P}$ +case] [$\bar{V}$] [I]
$\bar{P} \rightarrow N\ h$	$\bar{P} \rightarrow \bar{N}^{*}\ (P)$
[N]	[$\bar{N}$] [P]
$\bar{N} \rightarrow h$	$\bar{N} \rightarrow (N)$
	[N]

(1) Example phrase structure rules in basic and expanded form.

In the rule for I-bar, the first rule shown in (1), the P is taken to mean "any number of maximal projections of Postp", the V means "any number of maximal projections of Verb", and the h means I. That is, an I-bar consists of a sequence of zero or more P-bars followed by a sequence of zero or more V-bars, followed by at most one I.

The second part of the phrase structure rule is the set of *surface templates*. These specify the possible leading edges of the phrase type -- in particular, what the first one or two daughters are. For I-bar, the associated templates shown in (1) say that an I-bar may begin with a single P-bar (where P-bar is the maximal projection of P), or with a single V-bar. In addition, the set of templates includes the head template (implicit in all PS rules), which for I-bar says that an I-bar may begin with an I. The other rules shown are similar.

The phrase structure analyzer proceeds to describe an utterance by creating node variables, which are then described according to their categorial properties and their domination links to other nodes. The phrase structure component is the only module that makes assertions that some nodes *dominate* other nodes. Later components may add assertions (chiefly equality assertions between node names) that add information about which nodes dominate which others. We assume that the component which computes a semantic structure reinterprets this description in some fashion, perhaps by taking a circumscriptive interpretation of the domination predications, resulting in an interpretation which is equivalent to direct domination.

Rootedness and Intonational Phrases. In addition to the phrase structure analyzer, a prosodic module adds to the description of an utterance. We assume that every word in an intonational phrase is in the same syntactic phrase: put another way, the words of an intonational phrase are *rooted*. What that root phrase is and what its subconstituents are is the phrase structure's business. But even in the absence of any information about the phrase structure, the words of an intonational phrase will have a common root. Stated differently: there is a universal constituent analysis of any sentence, imposed by the intonational structure, and produced without reference to the phrase structure rules.

Interpreting Nodes: Post Phrase Structure. Phrase structure *nodes* are the entities in terms of which the phrase structure is described. However, these nodes, and the node variables that name them, do not end their active life there: they play a crucial role in the functioning of the several other components that are also busy describing the sentence. In particular, the nodes that are described by the phrase structure component are exactly the nodes that semantics uses to describe semantic structure, just as they are the nodes that are used to describe informational structure. Thus, according to this view, what we have is not a *mapping* from component to component, but rather an *incremental description* of the utterance. Viewed from the perspective of the post phrase er and the intonational analyzer -- is to provide node var(just like are lexical analyzer and the intonational analyzer) is to provide node variables, which are partially described, that the semantics can then describe more fully.

Although the various components use the same node variables, they don't necessarily use the same terms of description. Indeed, we assume that each component will have its own set of predicates and will follow its own internal constraints. Although this talk will offer only a rough picture of the constraints on the various components outside of the phrase structure, some of the properties of the relationships among the phrase structure and further processing are clear:

1) All nodes variables are created either by the phrase structure analyzer or by prior processes: namely, the lexical analyzer, the intonational analyzer, and the rootedness component. It is this fact that justifies calling semantic and informational processing *post* phrase structure. The moves of the phrase structure analyzer are logically prior because it provides the variables for the descriptions of the other components.

2) While the node variables that the phrase structure uses may corefer, the parser doesn't make use of this fact or even know of its possibility. All predication of equality among

nodes is done in later processes.

3) No predications that the semantics (that is, the post-phrase-structure) makes are necessary for further phrase structure processing.

4) Extracategorial items are described further by components after the phrase structure.

The Extracategorial Elements "wa" and "no". Given this framework we will argue that the particles "wa" and "no", as well as intonational breaks, are all extracategorial (though they of course differ in other respects). In parsing, these extracategorial elements have the effect of breaking up the description that the phrase structure analyzer is producing, since the phrase structure analyzer knows nothing about them. But it turns out that this results in a partial description that is exactly what is needed for the later modules to further specify: ambiguities are left open until the correct decision can be made, and the process of describing a sentence can proceed deterministically.

Technical Prospects of Telephone Translation

Makoto Nagao

Kyoto University

1. Basic and applied research

In order to develop a telephone conversation translation system, applied research on dialog over limited domains with restricted expressions must be pursued in parallel with basic research on dialog over unlimited domains. It will also be important to collect data on conversation over various topics carried out between Japanese and English speakers.

2. Hearer's model

Because, in uttering questions and declarations, the speaker is estimating the effect it will cause on the hearer, the system must include a dynamic model of the hearer's mentality, knowledge, topic continuity, plan, etc. Processing of utterances must be conducted on the basis of this model. The response also must be on the basis of this model.

For example, if a speaker's declaration of a fact contradicts the hearer's knowledge of the speaker, then the hearer

(i) asks the speaker's intention,
(ii) checks which one is true,
(iii) guesses that the speaker is not reliable,
etc.

If it does not contradict the hearer's knowledge,

(i) the hearer adds the fact to his world knowledge, and
(ii) the speaker adds the fact to his world knowledge of the hearer's model of the speaker.

3. System components

A telephone conversation translation system must solve the following problems:

(1) how to reflect speech features specific to the speakers

(2) how to recognize the domain of the conversation

(3) translation in consideration of the cultural background specific to the language, such as honorific expressions

(4) translation of imperfect, ungrammatical, and fragmentary expressons

(5) segmentation of utterances and construction of discourse structure

Particularly, in human conversation the hearer understands what was said by the speaker only partially and incompletely. Even misunderstanding is inevitable. If human conversation is to be exactly modeled, this will cause serious problems in developing such a system.

(Summarized by K. Yoshimoto)

PROGRAM

Workshop on Natural Language Dialogue Interpretation

FRIDAY, NOVEMBER 27

OPENING SESSION Chair: Teruaki Aizawa (ATR)

9:30-9:45 Kohei Habara (ATR International)
Opening Address

9:45-10:20 Akira Kurematsu (ATR)
Keynote Speech: Problems Involved in Telephone Interpretation.

10:20-10:50 COFFEE BREAK

SESSION 1 Chair: Shuji Doshita (Kyoto University)

10:50-11:25 Hitoshi Iida (ATR)
Distinctive Features of Conversations and Inter-keyboard Interpretation.

11:25-12:00 Jun-ichi Tsujii (Kyoto University)
Dialogue Translation vs Text Translation: Interpretation-Based Approach.

12:00-2:00 LUNCH

SESSION 2 Chair: Masaaki Yamanashi (Kyoto University)

2:00-2:35 Tadahiro Kitahashi (Osaka University)
An Application of Pragmatic Knowledge to a Bottom-up Analysis of Japanese Sentences.

2:35-3:20 C. Raymond Perrault (SRI International)
The Extension of Speech Act Theory to Multiple Sentences.

3:20-3:50 COFFEE BREAK

SESSION 3 Chair: Takao Gunji (Osaka University)

3:50-4:35 Jaime G. Carbonell and Masaru Tomita(CMU)
Knowledge-Based Machine Translation.

4:35-5:10 Naoyuki Okada (Oita University)
The System of Intellect and Emotion: An Engineering Approach to the Model of Mind.

5:10-5:45 Hiroaki Tsurumaru (Nagasaki University)
Theasaurus Construction from a Japanese Dictionary.

6:00-7:30 RECEPTION

SATURDAY, NOVEMBER 28

SESSION 4 Chair: Hirosato Nomura (NTT)

9:00-9:45 Julia Hirschberg (AT&T Bell Laboratories)
Intonational Meaning in the Interpretation of Discourse.

9:45-10:30 Donald Hindle (AT&T Bell Laboratories)
Parsing Natural Language Dialogue.

10:30-11:00 COFFEE BREAK

SESSION 5 Chair: Sho Yoshida (Kyushu Institute of Technology)

11:00-11:35 Makoto Nagao (Kyoto University)
Technical Prospects of Telephone Translation.

11:35-12:00 GENERAL DISCUSSION