TR-I-0128

Known Current Problems In Automatic Interpretation: Challenges for Language Understanding
自動翻訳に向けた対話理解の問題

John K. Myers and Takashi Toyoshima
真龍主·星音　豊島·孝之

January 19, 1990

Abstract

This paper provides an examination of some of the known current problems in the Interpreting Telephony task Such an examination is necessary in order to determine where natural language understanding efforts (i.e., plan recognition, dialog understanding, etc.) should be focussed. Examining the problems also provides explicit recognition of some of the assumptions that have been made in the design of the automatic interpretation architecture; these assumptions can then either be confirmed, or reevaluated and modified.

The results were obtained by analyzing ATR's "conversations 1-10" corpus, which is a current research target for automatic translation efforts. Thus, the results reflect the immediately relevant problems for matchine translation. Although undoubtedly more problems will be discovered as the state-of-the-art advances, this study has attempted to be a complete listing of the problems in the current corpus.

The paper first reviews an assumed architecture for automatic interpretation. Next, the known outstanding problems, as derived from the corpus, are presented in order of descending frequency, and are also tabulated in the appendix. The frequency is presented as a rough measure of the importance of the problem. The paper concludes with a brief discussion.

KNOWN CURRENT PROBLEMS IN AUTOMATIC INTERPRETATION: CHALLENGES FOR DIALOG UNDERSTANDING

John K. Myers and Takashi Toyoshima

ATR Interpreting Telephony Research Laboratories
Sanpeidani, Inuidani
Seika-cho, Soraku-gun, Kyoto 619-02, Japan
myers%atr-la.atr.junet@uunet.uu.net

Abstract

This paper provides an examination of some of the known problems in the Interpreting Telephony task that are not yet obviously solved. Such an examination is necessary in order to determine where natural language understanding efforts (i.e., plan recognition, dialog understanding, etc.) should be focussed. Examining the problems also provides explicit recognition of some of the assumptions that have been made in the design of the automatic interpretation architecture; these assumptions can then either be confirmed, or reevaluated and modified.

The results were obtained by analyzing ATR's "conversations 1-10" corpus, which is a current research target for automatic translation efforts. Thus, the results reflect the immediately relevant problems for machine translation. Undoubtedly, more problems will be discovered as the state-of-the-art advances.

The paper first reviews an assumed architecture for automatic interpretation. Next, the known outstanding problems, as derived from the corpus, are presented in order of descending frequency, and are also tabulated in the appendix. The frequency is presented as a rough measure of the importance of the problem. The paper concludes with a brief discussion.[1]

Contents

1 Introduction **2**

2 Assumed Architecture **2**

3 Some Known Outstanding Problems: Problems Ordered by Frequency **4**

3.1 Disambiguation of possible understood utterances 4
3.2 Prediction of Possible Subsequent Utterances 5
3.3 Prosody 6
3.4 Article Generation 6
3.5 Difficult Prepositions, Postpositions and Particles 7
3.6 Subject Generation 8
3.7 Verbal Honorifics and Humble Forms 9
3.8 Softening Moderations 9
3.9 Ill-Formed Input Utterances 10
3.10 Nominal Honorifics 11
3.11 "Will" Future or Commitment Form Generation Required 12
3.12 Incomplete Specification 12
3.13 ZERO Indirect Referent 13
3.14 Closing Signals 13
3.15 Plural Marker Generation Required 14
3.16 Ambiguous or Vague Vocabulary 14
3.17 Usage of Indefinites 15
3.18 Aspect Generation 16
3.19 Short Answers 17
3.20 Causative Transfer Problem 17
3.21 Difficult or Impossible Interpretations 18
3.22 Zero Verb and Case Markers 18
3.23 Gender Determination for Titles 19

4 Discussion **21**

5 Summary **21**

6 Conclusion **21**

A Results **24**

1 Introduction

This study attempts to provide a brief examination of some of the immediate problems facing automatic interpretation in general. From this, the most important challenges facing the plan recognition/natural language understanding system in particular can be perceived.

A particular architecture for the automatic interpretation system is assumed. Also, a particular input corpus, "conversations 1-10", is assumed. An implementation of the architecture does not currently exist; it is a future target. However, given this architecture, and given this input corpus, where will the system have problems in interpretation in general and in natural language understanding in particular?

A number of other works have discussed the various different problems described here (e.g., most notably [Iid88]). This work is distinct in that (1) it examines the translation requirements of Dialogs 1-10, which are a current research target to be solved in the coming three years; (2) it attempts an *exhaustive* listing of the problems encountered in this corpus; (3) it classifies the problems by frequency, which helps to indicate relative importance; (4) tentative recommendations are made as to which modules should be responsible for handling the problem.

2 Assumed Architecture

This section briefly reviews the main concepts of the machine translation architecture proposed and created by ATR Interpreting Telephony Research Laboratories. The assumed architecture is a target design only, and has not been assembled yet. The architecture is discussed more fully in [IKYA89], [KU89], and [MOAK89].

One user speaks into a telephone. A continuous-speech recognition system recognizes the phonemes in the utterance, and constructs words out of the phonemes. The speech-recognition system uses predicted possible utterance sentences to help recognize the phonemes and words. These predictions are generated at a high level by a conversation-understanding module, and at a low level by a NL generation system.

The result of the speech recognition system is a lattice of ordered possible words. Each path in the lattice has an assigned ranking or probability. The speech recognition system also returns a representation of the prosody associated with the sentence.

These results are sent to a parsing system, which performs syntactic and semantic parses on the possible sentences in the lattice. The parsing system rejects utterance parses that are seriously syntactically or semantically ill-formed, while accepting and representing parses for ill-formed utterances that the user actually says.

The multiple results of this parse are sent to the understanding system. The understanding system must use context, prosody, pragmatics, and other sources of information to disambiguate the parses and to fill in required missing information. The understanding system should result in a single structure to be translated.

The result of the understanding system is sent to a transfer module, which translates language-dependent concepts from the source language into equivalent concepts in the target language. The transfer module must also translate the prosody of the sentence.

From here, the sentence is sent to the natural-language generation system, which is

responsible for generating the syntax and the surface form of the utterance. When the generation system needs any information that has not been specified yet for decisions, it must ask the understanding system to supply this. The generation system also generates a representation for the surface intonation of the output.

Finally, a speech generation system takes the unambiguous instructions of the natural-language generation system, and creates speech sounds corresponding to the translated utterances. These are fed to the second user. In this manner, the system performs automatic interpretation.

3 Some Known Outstanding Problems: Problems Ordered by Frequency

Each known problem is discussed in one of the following subsections. First, the problem is defined, and an example is given. Next, the problem is discussed. Finally, tentative assumptions are made as to which module will be responsible for handling that problem. The problems are presented in their order of frequency of occurrence in the corpus. The frequency is presented as a rough measure of the importance of the problem.

Frequency is defined as the count of problem occurrence instances divided by the number of utterances. There were 224 utterances in the examined corpus.

It is significant that the corpus examined was a written corpus. Live spoken dialogs will introduce additional dynamic effects into the interpretation problem. In addition, it will then no longer be possible to recognize words based on kanji, and homophones will have to be considered.

3.1 Disambiguation of possible understood utterances

Frequency: more than 100% (estimated), according to the assumed architecture.

Definition: The continuous speech recognition process, the parsing process, and the illocutionary force understanding module all produce outputs consisting of multiple possibilities. Although many these possibilities can be eliminated by syntactic (e.g. [HY88]) or semantic constraints, a recent study at CMU showed that a single spoken utterance could yield around 60 different possible interpretations, all semantically well-formed [TT88]. The automatic translation system must disambiguate between different possibilities, and produce an output utterance that captures the actual meaning of the input utterance.

Example:

```
Do you have one?
```

("You" could be singular or plural; "one" could refer to any of several previously mentioned noun phrases).

Discussion: Disambiguation involves selecting the possibility that "makes the most sense" in the conversation, and discarding the others. This implies at least two parts: an evidential reasoning framework that can calculate the most probable meaning using supplied information, and a set of knowledge sources tuned towards whether an utterance "makes sense" or not, that can contribute their information to the disambiguation process. Such knowledge sources traditionally are taken from all levels of natural language understanding, including discourse understanding, pragmatics, domain knowledge, speech act analysis, and "common-sense understanding". It is still unknown just which knowledge sources are the most important and which contribute the most to everyday conversation. It is possible that many of these knowledge sources are not fully required for automatic interpretation.

The problem is made slightly more complex by the fact that some utterances have multiple legitimate meanings. For instance, "Can I leave the hotel reservation to you?" is both an ability question ("Am I able to have you take care of my hotel reservation?") and a polite request ("Please take care of it."). In many cases, such multiple meanings can be translated by condensing them back down into a single structure during generation, by using a similar multiple-meaning expression in the target language. However, in some cases generation will simply have to generate more than one utterance in the target language, to account for all the meaning in the source language.

A separate problem is how the disambiguation process will be able to represent and recognize possible legitimate multiple meanings, while still discarding incorrect possibilities. One problem here is that some utterances can have multiple illocutionary forces; thus, the system must be able to represent both cases where a sentence can mean one thing or another or both, and also cases where a sentence can mean one thing or another, but not both.

An additional difficulty for disambiguation will be the eventual problem of accepting the meaning of ill-formed utterances. This is dealt with separately under section 3.9.

Responsible Modules: It is assumed that the semantics will reject all possibilities that are (incorrectly) semantically ill-formed, or at least give them extremely low preferences. The pragmatics module should also reject all possible utterance meanings that are pragmatically ill-formed. With this as a given, the disambiguation task is important enough that a separate module inside the understanding system is proposed to deal with it. This module will probably take input from discourse understanding, domain knowledge, and intention/speech-act understanding. It might also be necessary to use partial results from the transfer module.

3.2 Prediction of Possible Subsequent Utterances

Frequency: 100% (estimated), according to the assumed architecture.

Definition: In order for the continuous-speech understanding system to work well, the other modules should generate predictions of the possible expected subsequent utterances. These should be sent to the speech understanding system at the word level.

Example:

```
Iie, mada desu.
PREDICTED NEXT UTTERANCE:  Wakarimashita.
```

Discussion: Prediction results can be used for disambiguation of the next utterance. If the predictions are ranked, possible input utterances that match predictions with high rankings should receive high likelihood scores.

One of the outstanding problems associated with using predictions as input to speech recognition is that the current level of predictions is entirely at the *semantic* or even *speech act* level, whereas the speech recognition module will probably require

predictions at the *lexical* level. This apparently requires generation of multiple possible utterances These can only be probabilistic recommendations.

One approach to prediction is to use a speech act grammar that suggests what type of utterance will most probably follow [AI88]. In this case, although the words themselves are not predicted, at least the general style or type of the sentence (whether it is a question, order, statement, or answer) can be predicted.

Responsible Modules: This problem deals with multiple utterances and is an understanding system/plan recognition problem.

3.3 Prosody

Frequency: 100% (estimated), according to the assumed architecture.

Definition: Prosody has to be able to be represented. Prosody must be parsed; understood; transferred; and generated.

Example:

```
Wakarimashita.
Wakarimashita...
Wakarimashita!
Wakarimashita?
Wakarimashita!!!
```

Discussion: Prosody analysis is most important when analyzing the speech-act intentions of an utterance. As the example shows, it is possible for the same lexical utterance to have widely different meanings. Other common utterances can also take on subtle shades of meaning, depending on the prosody.

Prosody may also be useful in syntactic and semantic parsing.

In [Iid88], Iida points out that prosody can be used to: (1) understand illocutionary acts; (2) help determine scoping in syntactic and semantic parsing; (3) discriminate previous information from newly introduced information, including referents; (4) recognize discourse markers; (5) recognize the illocutionary force of answers; and, (6) disambiguate demonstratives from interjections or hesitation forms in Japanese.

Responsible Modules: All modules must be able to represent and work with prosody information. Parsing must be able to parse low-level prosody information into high-level information; transfer and generation must be able to translate and generate appropriate prosody information. The understanding module must be able to use prosody to determine the underlying illocutionary and perlocutionary forces of an utterance, and to suggest corresponding appropriate prosody for the target language.

3.4 Article Generation

Frequency: 89%

Definition: Articles must be generated for nouns in English. An article can either be definite (the), indefinite (a, an, some), or zero (). Articles are not generated for nouns that are marked as objects of possessives or demonstratives, etc.

The problem of indefinites is treated separately.

Example:

```
Is there [a] discount for [] members?
```

Discussion: Japanese does not seem to use the information contained in English articles. Thus, probably in most cases this information can simply be dropped when generating Japanese from English, and there will be no problem. In some cases, as a refinement, the article could be interpreted as a demonstrative ("kono", "sono", "ano"); this would require basic research as to when it should be performed or omitted. However, such a refinement can be ignored for the immediate future.

Article determination requires that the corresponding noun's referent be positively identified, as discussed in [NI89].

Responsible Modules: The discourse model and the conversation history will probably contribute information to help solve this problem, along with the knowledge base. The natural language generation module will probably be mainly responsible for article generation.

3.5 Difficult Prepositions, Postpositions and Particles

Frequency: 28%

Definition: English prepositions, Japanese postpositions, and particles from both languages contribute many important shades of meaning but can be extremely difficult to interpret.

Example:

```
Nihon-go no hon
([a] Japanese book)
([a] book of the Japanese language OR The Japanese language's book)
([a] book on [the topic of] Japanese)
([a] book [written] in Japanese)

ryou no waribiki
([a] discount of [the] fee)
([a] discount for [the] fee)
([a] discount in [the] fee)
```

Discussion: Particles are widespread, and often cause problems. The difficulty in interpretation can lie mostly with the source language, when a particular particle has different ambiguous meanings. Alternatively, it can lie mostly with the target language, when an unambiguous but vague meaning can be transferred into different interpretations.

The way that this problem is handled depends heavily on the design of the system. If a mostly language-dependent parsing strategy is used, meanings from unambiguous but vague particles may not be separated until the generation system cannot translate them. If a mostly language-independent parsing strategy is used, the parser might detect and generate all the different possible interpretations, requiring the understanding system to disambiguate them early on.

This problem also includes the two English conditionals "if" and "when".

Responsible Modules: Parsing and generation must flag when a particle phrase can have different interpretations that cause problems in generation. The understanding system must disambiguate the specified alternative interpretations.

3.6 Subject Generation

Frequency: 25%

Definition: English sentences require that a subject be supplied. The subject depends on the verb and the conceptualization of the action or situation being described. However, often in Japanese sentences the noun phrase corresponding to the English subject may not be mentioned, or it may be difficult to determine.

The interpreting system must be able to generate the subject for English sentences when the Japanese sentence has a zero topic or when the subject is otherwise difficult to determine. Also, the system should be able to determine when a topic should be underspecified into a zero topic during the generation of Japanese sentences [Yos88].

Sentences that could be disambiguated based on honorific information were not included in this frequency score.

Example:

```
Musume wa kaigi niwa sanka shinai no desu ga.
(lit.: Concerning daughter, concerning at the conference,
   someone will not attend...)
([My] daughter [will] not attend the conference...)
Kanko dake sanka dekimasu desho ka?
([She] is probably able to participate in only the tour?)
Ima Kyouto-eki ni iru n desu.
(Now [I] am at Kyoto station.)
Kyouto Hoteru to Kyouto Purinsu Hoteru wa yoyaku dekimasu.
([We] can reserve rooms [for you] at the Kyoto Hotel and the Kyoto Prince Hotel.)
```

Discussion: This is a difficult problem, and one that must usually be solved by plan recognition. Dialog understanding should also provide some information. As has been mentioned, honorific information has been used to disambiguate subjects [Yos88], and problems resolvable by this method were not counted here.

Besides simple sentence subject determination, there is also the problem of determining the subjects of subordinate clauses.

In limited dialogs, it is possible to assume that there is a high prior probability that the missing subject is either the speaker or the hearer. Although this heuristic would provide useful input to an evidential reasoning system, it cannot successfully be used as a true premise in realistic conversations.

In addition, generation of natural Japanese sentences where the natural topic is not the English sentence subject appears to be an unsolved research problem.

Responsible Modules: Although some subjects can be determined by the parsing or the generation systems, for the most part this is a dialog understanding system problem.

3.7 Verbal Honorifics and Humble Forms

Frequency: 21%

Definition: Japanese uses verbal honorific strategies when the speaker talks about an action of someone else and wants to show respect or politeness. Verbal humble strategies are used when the speaker talks about his own action and wants to show humbleness or politeness. These strategies must be appropriately generated when interpreting English into Japanese. A similar but appropriate politeness, respect, and/or humbleness attitude should be generated when interpreting Japanese into English, as is pointed out in [KU89].

Example:

```
Touroku-youshi o o-okuri-itashimasu.
```

Discussion: The main difficulty with this problem seems to be understanding when and how to generate the honorific strategies. Since English basically does not use the equivalent markings, the required information must be drawn from situation-specific details.

Responsible Modules: This seems to be clearly a transfer problem. The transfer module will probably use information from discourse to help determine when it is appropriate to generate honorifics.

3.8 Softening Moderations

Frequency: 16%

Definition: Japanese has a few strategies for "softening" the impact of an utterance. The most important of these are the (presumptive) "deshou" and the "no desu ga" copulas. Equivalent English strategies are starting a sentence with "Maybe" ("Probably", "Perhaps", etc.) or with the oral hesitation form "Um,", or ending a sentence with a trailing intonation (represented by an ellipsis "..."). However, English does not use softeners nearly as much as Japanese does. Thus, the primary problem with softening is deciding whether to transfer the moderation as is or to drop it (when interpreting into English), or deciding when to introduce the softening moderation (when interpreting into Japanese). A secondary problem is determining the proper and appropriate translation form.

Example:

```
Kaigi ni ronbun o happyou-shitai to omotte-iru no desu ga.
```

Discussion: The softeners are unambiguous but vague. More research is required in the uses of vague language. More research is also required to determine the exact type of information encoded by the softening moderation forms.

Iida notes the need for understanding softeners in [Iid88].

Responsible Modules: This seems to be a transfer problem. It is also possible that softeners are used to indicate indirect messages, in which case the illocutionary force recognition part of the understanding system would use these as input.

3.9 Ill-Formed Input Utterances

Frequency: 0% in the employed corpus. (Estimated around 15% from a small spoken corpus.)

Definition: An input utterance is ill-formed when it does not constitute a syntactically well-formed sentence or phrase. This includes false starts, "uh" noises in the middle of sentences, ungrammatical spoken language, loose verbal case usage, and utterances where the speaker changes his train of thought in the middle of a sentence. (These subproblems are distinct, and should actually be dealt with separately when the time comes.)

Example: (taken from a spoken corpus)

```
  Yes, I would, if you don't mind, would you
please send me the two sets of the registration form, can you?

Yes, certainly uh, by the way, if you uh, if you are rushed to finish
our summary, you could also fax it to us...

Uh, let me send you uh, the list of our members and in
addition to your sending them invitations, please send me
not invitations, announcements, excuse me.
```

Discussion: It is important to use native speaker data when analyzing this problem. A non-native-speaker human interpreter may use more ill-formed utterances than a native-speaker conversation participant. Also, it may be that even a native-speaker interpreter might use slightly more ill-formed utterances than normal, due to the pressure of having to interpret. Frequency of usage will also probably vary with education and social status.

An outstanding problem that will not be considered until later in the future is the question of how to interpret and generate the interpreted language resulting from such utterances. What kind of information is communicated by the different types of ill-formed utterances (e.g. ignorance, uncertainty, politeness)? Should this information be preserved, or should the utterance be "cleaned up"? Will an interpreting telephone user feel more comfortable if an occasional ill-formed utterance is generated?

One particular type of problem is pointed out by Iida [Iid88], who calls attention to the fact that Japanese spoken sentences can sometimes drop postpositions (usually case markers such as "ga", "ni", or "o").

Kudo et al. also describe a method for accepting ill-formed English written input from non-native speakers [KKCM88].

Responsible Modules: The syntactic and semantic parsers will have to be able to accept such actual ill-formed utterances, while rejecting or giving low ratings to "obviously bad" incorrect possible ill-formed utterances proposed by the speech recognition process. Other modules will have to be able to represent and work with the concepts communicated in the utterance. Transfer may have to interpret the ill-formedness into an analogous format for the target language, and generation may have to be able to generate such utterances.

3.10 Nominal Honorifics

Frequency: 14%

Definition: Japanese uses the nominal honorific prefixes "o-" or "go-" when the speaker talks about a noun in the personal sphere of a different person and wants to be polite, but not when the speaker talks about nouns in his or her own personal sphere. These prefixes must be generated when interpreting English into Japanese. If possible, a similar but appropriate politeness attitude should be generated when interpreting Japanese into English.

Example:

```
Go-juusho to o-namae o onegaishimasu.
```

Discussion: The main problem here again is generating the Japanese. Additional situation-dependent information is required from the discourse model and the understanding system.

Responsible Modules: This is a disambiguation problem. It can be handled partially by the semantic parser and the natural-language generator; however, these must ask the understanding system for help when the required information cannot otherwise be determined. Deciding when to generate nominal honorifics seems to be the responsibility of the transfer module.

3.11 "Will" Future or Commitment Form Generation Required

Frequency: 12%

Definition: The "will", "shall", or "am going to" form in English is variously believed to be a future tense marker, a volitional/commitment modal marker, or both. Japanese typically does not use an analogous form. Thus, this form must be generated when interpreting a Japanese future or commitment sentence into English.

Example:

```
Kaigi no anai-sho o o-okuri-itashimasu
(lit., I send you the conference announcement)
(I [will] send you the conference announcement)
```

Discussion: This is a difficult problem that rests on determining when a speaker is talking about something in the future tense or is making a commitment. Pragmatic and discourse information sources must be used by the understanding system to make this determination.

Responsible Modules: Although semantic parsing can sometimes help, this seems to be an understanding system problem.

3.12 Incomplete Specification

Frequency: 12%

Definition: Japanese allows referentials to be less specific than is permitted in English. This happens especially in the case of possessives, when something is owned or is part of someone.

The subsumed problem of indefinites is treated separately, as is the case of short answers.

Example:

```
Go-juusho to o-namae onegaishimasu.
(lit.: [A] name and [an] address, please.)
(Your name and address, please.)
```

Discussion: Incomplete specification is actually another strategy for being indirect, and as such carries politeness/honorific connotations.

The amount of specification required in a reference is language-dependent, and is usually chosen by the generation system. However, the transfer system might specify this in stereotypical situations. The transfer or generation system must notify the understanding system when a reference is underspecified and more information is required to complete it.

Underspecifying a reference is much easier to perform and is a generation problem. However, when to underspecify is still a research issue.

Responsible Modules: Discourse and pragmatic information will probably be used by the understanding system to disambiguate and further specify references when needed. The transfer and generation systems must indicate when further specification is required.

3.13 ZERO Indirect Referent

Frequency: 12%

Definition: Japanese allows the indirect object in dative verb phrases and other ditransitive verbs (those requiring an indirect referent) to be elided. English syntax usually requires the indirect referent to be supplied. Thus, the object must be generated when interpreting into English. Likewise, the decision as to when to elide the object must be provided when interpreting into Japanese.

Example:

```
Kaigi ni tsuite kuwashii koto o oshiete-kudasai.
(Please tell [me] about the details of the conference.
```

Discussion: Ditransitive verbs are quite common, and include words such as ask, tell, give, send, etc. The problem can depend on the word-choice used for translation: certain vocabulary words in the target language can sometimes encode the same concept, while not requiring an indirect referent.

Sometimes the required information can be obtained through parsing and the use of pragmatic constraints, especially when a humble form is being used. However, often the information will simply have to be determined by dialog understanding. Script and/or plan recognition may help in this situation as well.

Responsible Modules: Parsing will solve some problems. Generation will solve a few problems. Generation must flag when an unsolved problem is occurring and ask the understanding system to supply the required information.

3.14 Closing Signals

Frequency: 7%

Definition: The Japanese apologies "shitsurei-shimasu" and occasionally "sumimasen" can also be used to signal initiation of conversation closure. In this case they must not be interpreted as apologies when generating English, but must be transferred into similar closing signals.

Example:

```
Shitsurei-shimasu.
(lit.:  I am doing [something] rude)
(usual: I'm sorry)
(at conv. close: Well, take care, then.)
```

Discussion: The main problem here is detecting when a conversation is *about* to end, or when a person should be making an apology. It is thus an illocutionary force problem, which will probably use at least plan recognition methods.

Responsible Modules: This is probably the responsibility of the plan recognition module. Discourse information will also be useful.

3.15 Plural Marker Generation Required

Frequency: 6%

Definition: Japanese has an extremely limited plural system, and in almost all cases plural nouns are unmarked. Verbs are never marked. However, English must mark plural nouns and corresponding verbs. It is necessary for the interpreting system to recognize unmarked plural nouns in Japanese, and generate the plural marking.

Example:

```
Kouen-sha mo sanka-sareru no desu ka?
(lit., Is [the] speaker also participating?)
(Are [the] speakers also participating?)
```

Discussion: This is in general a hard problem, and can usually only be solved by world knowledge, common-sense knowledge, or probabilistic assumptions, if then. There is typically no semantic way of telling when a noun is actually plural in Japanese. However, this is a very strong marking in English, and mis-interpretation leads to objectionable results.

Responsible Modules: The parsing system must flag when a noun or noun phrase could be plural. The understanding system as a whole must attempt to disambiguate whether a noun is singular or plural.

3.16 Ambiguous or Vague Vocabulary

Frequency: 5%

Definition: The problem of ambiguous or vague vocabulary occurs when a word in the source language translates into two or more words in the target language with significantly different meanings. This problem occurs both when the target language is English and when it is Japanese.

Example:

```
Kaigi ni tsuite kuwashii koto o oshiete-kudasai.
(Please teach [me] in detail the things about the conference.)
(Please tell [me] all about the details of the conference.)
```

Discussion: Perhaps one of the most important vague words in Japanese is the verb "onegaishimasu" ("–, please")("please 'take care of' –").

The obtained frequency result seems a bit low, perhaps because the input sentences were hand-parsed. The number of ambiguities will probably rise when a larger dictionary with multiple meanings is used by the computer.

Note that only problem vague words were counted. It is usually alright to translate a vague expression in one language into a vague expression in another language if a good equivalent translation can be found.

Responsible Modules: Some of this ambiguity can be taken care of by the parsing system. Some of it can be transferred and generated successfully. The remaining ambiguity and vagueness must be disambiguated by the understanding system.

3.17 Usage of Indefinites

Frequency: 4%

Definition: The indefinites are the English words "some" and "any", and also such words as "all (of the)" "every" "none (of the)", the pronoun "one", etc. They can be used by themselves as pronouns; with another noun as an indefinite adjective/article; or combined with "-one" or "-body", etc., as indefinite pronouns. The correct indefinite must be transferred into English from a corresponding form in Japanese, or generated if there is no corresponding form. When interpreting into Japanese, the English indefinite must be transferred appropriately.

This problem is thus closely linked with the problems of plural generation and article determination.

Example:

```
Dewa, dare-ka ga watashi no kawari ni sanka-suru koto wa dekimasu ka?
(Well then, can anybody participate instead of me?)
(Well then, can somebody participate instead of me?)
```

Discussion: The difference between "some" and "any" is subtle but important. "Any" usually implies an unrestricted or meta-restricted class, while "some" implies an indefinite member or members of an implicitly restricted class. The theory behind this is poor; for instance, is the class of "all researchers" implicitly restricted or meta-restricted? Until the theory behind class attitudes can be soundly defined, there is little hope of building a logical indefinite-usage module. Even after the theory is well-defined, there will still be the problem of recognizing when one nuance is meant and when the other is required.

Both "some" and "any" are used often. Note that they must be used with indefinite plural and mass nouns ("some information") instead of "a".

The problem of transferring English indefinites into Japanese appears to be easier, as information can be discarded.

Responsible Modules: This is a difficult problem that will have to be solved by an interaction between the parsing, understanding, transfer and generation modules.

3.18 Aspect Generation

Frequency: 2%

Definition: Japanese does not make use of the perfective aspect nearly as much as English does. However, the perfective conveys significant information in English. The interpreting system must recognize when the perfective aspect should be introduced when translating into English. Conversely, the system must decide whether it is more natural to drop the perfective or to transfer it when generating Japanese.

Example:

```
Eigo e no douji-tsuuyaku o youi-shite-imasu.
(We are preparing simultaneous interpretation into English.)
(We have prepared simultaneous interpretation into English.)
```

Discussion: The main problem here is determining when the perfective should be used. This requires understanding of the situation the speaker is talking about. Thus, world knowledge, common-sense knowledge, and plan recognition will have to be used to attack this problem. However, there still may be insufficient information in the conversation history to determine the proper use.

The theoretical problem of just exactly what information the perfective represents can be postponed for a long period, but will have to be attacked eventually in order to attain completely accurate interpretation.

Responsible Modules: This is a transfer problem, but the parsing module is also involved. The understanding system can help decide whether a sentence represents a perfective concept or not.

3.19 Short Answers

Frequency: 2%

Definition: In both Japanese and English, it is possible to provide short answers to yes/no questions by returning part of the question sentence, marked positively or negatively. Japanese returns the verb, whereas English returns the subject and the verbal auxiliary (but not the verb). Thus, when translating into Japanese, the elided verb must be provided; when translating into English, the elided subject and the auxiliary must be provided.

Example:

```
Kaigi no annai-sho wa o-mochi desu ka?
(About the conference announcement, [you] have?)
Motte-imasu.
(Have.)

Do you have a conference announcement?
I do.
```

Discussion: Recognizing when a short-answer sequence is occurring is a plan-recognition problem, as it deals with sequences of utterances.

Responsible Modules: Plan recognition.

3.20 Causative Transfer Problem

Frequency: 2%

Definition: Japanese uses the causative with verbs and in situations where English cannot. When interpreting Japanese into English, the meaning of the causative must be understood and transferred properly. When interpreting English into Japanese, appropriate causative situations must be recognized and the causative voice generated.

Example:

```
Touroku-youshi wa shikyuu okurasete-itadakimasu.
(lit.:  As for the registration form, [I] humbly-receive-the-favor-of
[you] making [me] send [it][to you].)
([I] [will] send [you] the registration form.)
```

Discussion: The Japanese language uses causatives because the conceptualization of actions and processes is quite different in some cases. It is important to recognize these cases and translate them properly, as literal translations are unacceptable in either case.

Responsible Modules: The transfer module must handle these cases.

3.21 Difficult or Impossible Interpretations

Frequency: 2%

Definition: Some concepts or phrases belong uniquely to usage situations inherent in the source culture. They have no corresponding meaning in the target culture. It can be difficult or almost impossible to interpret such phrases correctly.

Examples:

```
Yoroshiku onegaishimasu (lit.: Please think well [of me])
  {All-purpose polite request for favorable consideration,
   typically used after introductions and as a farewell greeting}

Irasshaimase!   (Come on in!)
  {Shouted by shopkeepers and cafeteria workers as a welcoming greeting}

kotatsu {A special coffee-table with heating coils on the underside
  and a quilted skirt}

tetsuzuki o suru (go through due formalities)(go through the proceedures of)
  (take steps in)   {A noun/verb meaning to take care of the red tape
  associated with accomplishing something e.g. in business or government}

Oops {Said by a person who has just dropped or broken something}

Have a nice day { Los Angeles general polite phrase, sometimes used as a
farewell greeting}
```

Discussion: Recognition of these phrases is generally not a problem; the only problem is in the actual translation. The economics of conversation time will prevent a full explanation. Since even people have difficulty coming up with good translations for these, whatever the computer can do should be acceptable.

An advanced system could be able to transfer some of these phrases by using case-based reasoning on analogous situations and illocutionary forces in the target language.

Responsible Modules: This is a transfer problem.

3.22 Zero Verb and Case Markers

Frequency: 1%

Definition: Japanese sometimes permits elision of the verb phrase and case particles, and replacement by the copula. The system must recognize such cases. The copula cannot be transferred directly, but must be replaced by the zero verb. In addition, the object of the copula must be transferred into the proper case in English. When interpreting English into Japanese, the system should recognize when such elision would sound natural, and perform it during transfer.

Example:

```
Watashi wa tempura o tabemasu.  Anata wa? / Watashi wa sushi desu.
(lit.:  I [will] eat tempura.  You? / I am [a] sushi.)
(I [will] [eat] sushi.)

Sanka-ryou wa ginkou-furikomi desu.
(lit.: [The] attendance fee is [a] bank-transfer.)
(The attendance fee is [paid] [by] bank-transfer.)
```

Discussion: Correct interpretation of this form of sentence can be broken up into two problems: (1) recognizing that the problem is occurring, and (2) supplying the elided verb and case markers.

This type of sentence is particularly difficult because often it will make semantic sense, and it can only be recognized by the difficulties encountered by pragmatics and the world knowledge module. If this form is typically used in certain types of situations, it might be possible to anticipate the possible usage of such a form; however, this would require further research.

The discourse module will probably be used to generate alternatives, and the disambiguation of the alternatives will have to be performed by the understanding system.

Responsible Modules: In some cases, this problem could be flagged by the semantic parse. In most cases, however, this will have to be noticed by the expectation part of the understanding system. The understanding system will have to generate and disambiguate the alternatives.

3.23 Gender Determination for Titles

Frequency: 1%

Definition: The Japanese title suffixes "-san" and "-sama" translate into the prefixes "Mr." or "Ms." (optionally "Mrs." or "Miss") depending on the gender of the referent.

Example:

```
Jinkou-Chinou-Kenkyuu-Jo no Jouji Ohara-sama desu ne.
([Mr.] George Ohara from AI Labs, right?)
```

Discussion: There are two straightforward ways of determining the gender of a name. If it is the name of one of the conversation participants, the speech recognition process should be able to distinguish gender in most cases. The other method is to have a table, indexed by gender, of the most common first names of both American and Japanese people, which can be used for determination.

Auxiliary verification can be provided by having the pragmatics module track the genders of third-person pronoun referents.

Another problem is what hedging strategies should be performed when the gender is unknown or ambiguous. Speech generation could possibly generate a sound halfway between "Mr." and "Ms.", or the language generation module could work around it in the sentence.

Responsible Modules: The illocutionary force module should recognize when a person is stating his or her own name. The speech recognition process should be able to identify the gender of a speaker. The world knowledge module in the understanding system should have a list of first names sorted by gender. Natural language generation should flag when a problem is occurring, and ask the understanding system to disambiguate.

4 Discussion

Many of the problems discussed here relate to translating English into Japanese. Since current primary efforts concern the translation of Japanese into English, a lesser emphasis may be placed on these problems in the near term.

It is important to note that these are only the problem found in the examined small corpus. Thus, the frequency data is skewed, when compared with conversations in general. Undoubtedly further problems will be found when more complex corpora begin to be processed.

5 Summary

It can be seen from this study that there are a number of significant problems facing an understanding/plan recognition module.

The most important problem appears to be disambiguating possible sentence parses. This requires an evidential reasoning module, to rank and evaluate the different possibilities. A representation for accepting multiple possible utterances from the parsing results will also have to be built.

The understanding module will probably also be responsible for generating the semantic form of predicted utterances for the speech recognition system. This will require interface protocols between the understanding and natural-language generation systems, again for passing multiple possible utterances.

The understanding module must also be able to represent and use prosody information for disambiguation and for indirect illocutionary force recognition.

The interplay between the understanding system and the transfer system, as well as the parsing and generation systems, must be explored and defined better. It is not clear what the requirements of each system are. In particular, the responsibility and the process for noticing problems, calling attention to them, and having the understanding system work on them, will have to be more clearly defined.

Finally, it is noted that many of the remaining outstanding problems, such as subject determination, "will" generation, and plural marker generation, can be cast in the form of disambiguation problems requiring evidential reasoning. This also tends to indicate that the disambiguation challenge is the most important problem to be attacked next.

6 Conclusion

A study of the actual current problems found in translating dialogs in the ATR "conversations 1-10" corpus has been presented. Although undoubtedly more problems will be found, the study has attempted to be exhaustive. General methods of attack and the interplay between responsible modules in the translation system have been proposed for each problem. The frequency of each problem encountered in the corpus has been presented as a rough measure of the importance of the problem. The resulting study demonstrates a series of challenges for an understanding system to attempt to handle,

gives a clearer view of what is required in order to perform good machine translation, and provides a rough plan as to where research efforts should next be allocated.

References

[AI88] Hidekazu Arita and Hitoshi Iida. Prediction of the next utterance in a task-oriented dialogue. In *IPSJ Fall Meeting*, 1988.

[Doh89] Kohji Dohsaka. Utterance interpretation based on constraints on dialogue paricipants' mental states. *Journal of Computer Software*, 6(4), October 1989.

[HY88] Tadasu Hattori and Kei Yoshimoto. Disambiguating japanese negative sentences. In *The 63rd Annual Meeting of the Linguistic Society of America*, 1988.

[Iid88] Hitoshi Iida. Pragmatic characteristics of natural spoken dialogues and dialogue processing issues. *Journal of JSAI*, 3(4), July 1988.

[IKYA89] Hitoshi Iida, Kiyoshi Kogure, Kei Yoshimoto, and Teruaki Aizawa. An experimental spoken natural dialogue translation system using a lexicon-driven grammar. In *Computer World 89 in Osaka*, 1989.

[KIYA89] Kiyoshi Kogure, Hitoshi Iida, Kei Yoshimoto, and Teruaki Aizawa. A new paradigm of dialogue translation. In *Computer World 89 in Osaka*, 1989.

[KKCM88] Ikuo Kudo, Hideya Koshino, Moonkyung Chung, and Tsuyoshi Morimoto. Schema method: a framework for correcting grammatically ill-formed input. In *12th International Conference on Computational Linguistics (COLING '88)*, 1988.

[KMY88] Masako Kume, Hiroyuki Maeda, and Kei Yoshimoto. Utilization of illocutionary force types for machine translation. In *IPSJ Fall Meeting*, 1988. (in Japanese).

[KSY89] Masako Kume, Gayle K. Sato, and Kei Yoshimoto. A descriptive framework for translating speaker's meaning. In *European Chapter of ACL89*, 1989.

[KU89] Akira Kurematsu and Yoshihiro Ueda. Generation in dialogue translation. In *Machine Translation Workshop at Univ of Manchester*, 1989.

[KY89] Masako Kume and Kei Yoshimoto. Pragmatic constraints on illocutionary force indicators. In *IPSJ Spring Meeting*, 1989. (in Japanese).

[MOAK89] Tsuyoshi Morimoto, Kentaro Ogura, Teruaki Aizawa, and Akira Kurematsu. Outline of an experimental spoken language translation system from japanese to english. In *IPSJ Fall Meeting*, 1989. (in Japanese).

[NI89] Izuru Nogaito and Hitoshi Iida. A method of semantic identification for noun phrases in dialogues and its application. In *IPSJ Spring Meeting*, 1989. (in Japanese).

[TT88] Hideto Tomabechi and Masaru Tomita. The integration of unification-based syntax/semantics and memory-based pragmatics for real-time understanding of noisy continuous speech input. In *AAAI'88: The Seventh National Conference on Artificial Intelligence*, pages 724–728, St. Paul, MN., 1988.

[Yos88] Kei Yoshimoto. Identifying zero pronouns in japanese dialogue. In *12th International Conference on Computational Linguistics (COLING '88)*, 1988.

A Results

PROBLEM	% UTTERANCES
Disambiguation of possible understood utterances	more than 100
Prediction	100
Prosody	100
Article Generation	89
Difficult Prepositions, Postpositions and Particles	28
Subject Generation	25
Verbal Honorifics and Humble Forms	21
Softening Aspect	16
Ill-Formed Input Utterances	15 (est.)
Nominal Honorifics	14
"Will" Future or Commitment Form Generation Required	12
Incomplete Specification	12
ZERO Indirect Referent	12
Closing Signals	7
Plural Generation Required	6
Ambiguous or Vague Vocabulary	5
Usage of Indefinites	4
Aspect Generation	2
Short Answers	2
Causative Transfer Problem	2
Difficult or Impossible Interpretations	2
Zero Verb and Case Markers	1
Gender Determination for Titles	1