Internal Use Only

002

TR-IT-0193

ATR-GETA Multimodal Interactive Disambiguation (MIDDIM) Collaborative Research Project Final Report

Kyung-ho Loken-Kim

1996, 9

ABSTRACT

A three year cooperative research project between ATR-ITL and GETA was conducted to investigate interactive disambiguation techniques of written and spoken input in a multimedia context. This paper first explains the scope of the project, second summarizes the joint research efforts made during the last three years in chronological order, and then, reports the outcome of the project.

ATR Interpreting Telecommunications Research Laboratories

Contents

1. SCOPE OF THE PROJECT

This paper will briefly summarize the 3-year cooperative research project (from July 1993 ~ summer of 1996) between Advanced Telecommunications Research International-Interpreting Telecommunication Research Laboratories (ATR-ITL) of Kyoto, Japan, and Study Group for Machine Translation (GETA) of Grenoble, France. This project was conducted as part of the 5 year cooperation on science and technology between ATR and the Centre National de la Recherche Scientifique (CNRS) which began in May, 1991. The objective of the project was to investigate the optimum combinations of interactive disambiguation modalities (techniques) that are adequate for written and spoken input in a multimedia context [1, P.6]. It was agreed that ATR-ITL should research interactive disambiguation of speech input while GETA would focus on text input.

Ambiguities are prevalent whether sentences are written or spoken. Ambiguity arises whenever there is more than one interpretation for a given expression. People apparently effortlessly resolve all sorts of ambiguities while continuing a conversation, reading books, etc.. There seem to exist two basic strategies for resolving ambiguities - guessing and questioning. Guessing requires an understanding of the context, experience, common sense, and world knowledge. Questioning is simpler, but interrupts the flow of the dialogue and reading. Frequent interruption is, at best, annoying, therefore, people rely on guessing for most ambiguities.

A computer system designed to help users resolve ambiguities, therefore, should be knowledge intensive so that automatic reduction of the ambiguities can be performed as much as possible [5, p.1]. Generating disambiguation questions should be the last resort. When questions are inevitable, they should be generated timely and clearly. On the other hand, users should actively participate in the disambiguation process by guiding (e.g., inserting disambiguation marks, and interrupting to correct the utterance being processed) and clarifying (e.g., answering clarification questions presented by the system, selecting among multiple intermediate results while the system is processing the utterance, or correcting some of them) [5, p.5] . Guidance and clarification - the process of interactively resolving ambiguities can be facilitated when a mixture of modalities are applied. In a multimedia setting, a spectrum of media is at our disposal for the presentation and generation of disambiguation questions. Disambiguation questions should be generated using the medium, or mixture of media that are best suited for resolving that kind of ambiguity. For example, presenting a textual explanation for graphic information would be possible but not practical [Winship, Minutes 2/28/94]. Answers to such questions can be multimodal. We can write, talk, and draw utilizing whatever media is made available to us.

2. MIDDIM AMBIGUITY DATABASE

The MIDDIM Ambiguity Database was created first, to assess the kinds of ambiguities arising in real dialogues and texts, and second, to device a means for effective interactive disambiguation [3, p.7]. The database contains 1) manually labeled ambiguities, as well as 2) manually and automatically generated ambiguity trees.

2.1. Definition of Ambiguity

As was mentioned before, ambiguity arises whenever there is more than one interpretation for a given expression. Sentences, such as, "Fred sees the woman in the park" and "John saw the man with the telescope" are well known examples of ambiguity (i.e., it is uncertain who is in the park and who has the telescope) [MIDDIM Minutes, 2/14/1994]. According to Boitet's [5, p. 4] definition of ambiguity, "a fragment F presents an ambiguity of degree n (n>=2) in an utterance U if it has n

different representations which can be extended in the same way to give a complete representation of U. To qualify as support of the ambiguity, F should further be minimal relative to that ambiguity, which means that F and its associated n representations cannot be reduced to a strictly smaller fragment F' and the n associated sub-representations without loosing the first property." For example, the fragment "*international telephone services*" in the sentence "Do you know where the *international telephone services* are located ?" with the two representations the "*[international telephone] services* " and the "*international [telephone services]*" is not minimal, therefore, it is ambiguous.

2.2. Classification of Ambiguity

Ambiguities occur in many different forms. According to the ATR-ITL-GETA ambiguity classification scheme [1], the above example "international telephone services" is classified as a segmentation ambiguity. In [1, pp. 9-11], there are six categories of ambiguities: 1) segmentation, 2) actualization, 3) communication, 4) modality, 5) reference, and 6) intention. Segmentation ambiguity, for instance, occurs when there is more than one way of segmenting a fragment. This happens "in

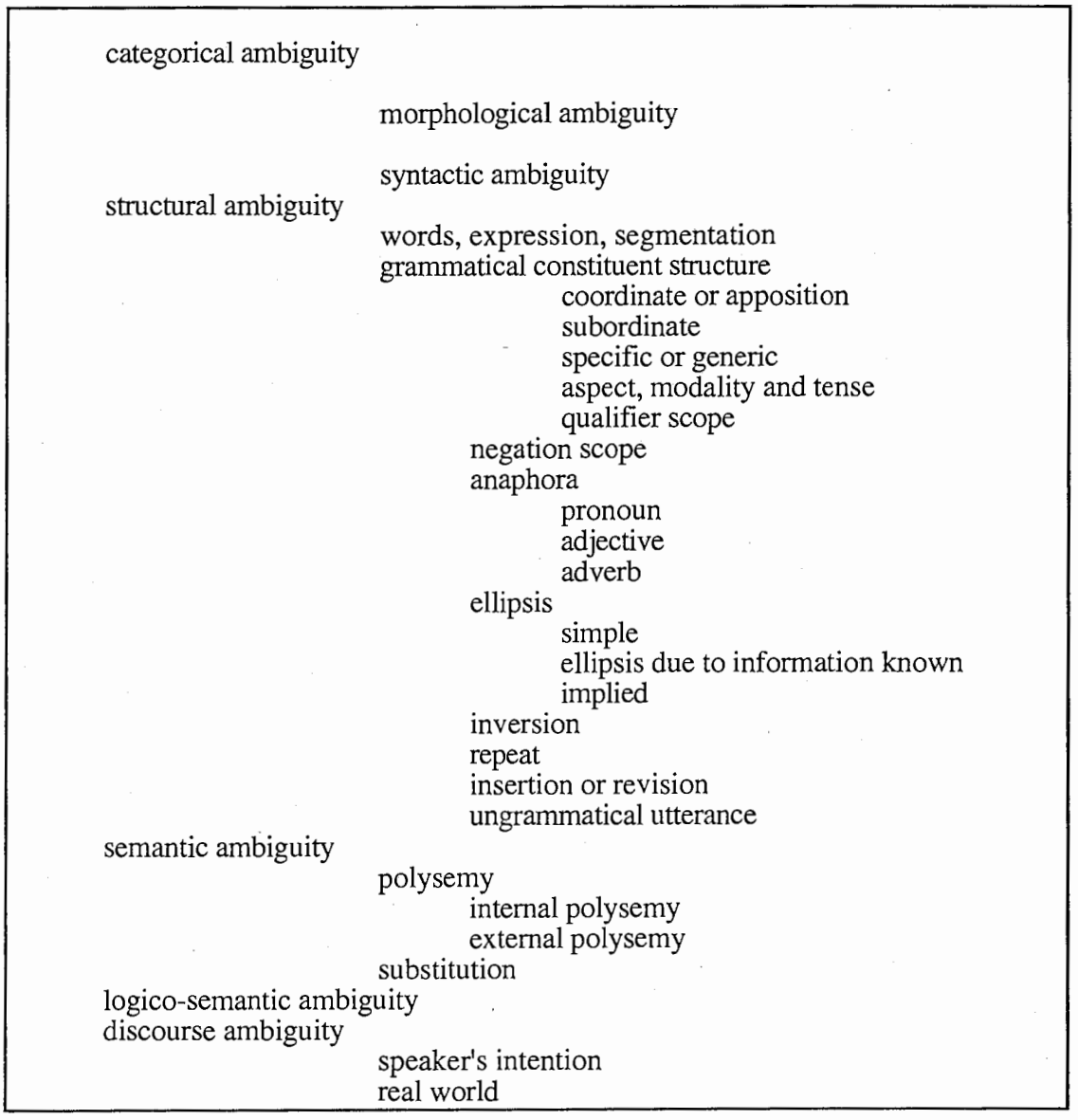


Figure 1. Ambiguity classification

transcribing dialogue from sounds, and also within terms and sentences of written text" [MIDDIM minutes, 2/28/94]. Other ambiguities, such as, ambiguity in intention "implied by words that can be different in different circumstances, referential ambiguity and so on" [MIDDIM minutes, 2/28/94] are all too familiar to us.

Further refinement of the classification lead to the one illustrated in Figure 1 [3, pp. 9~16], and it was used in manually labeling ambiguities to create the MIDDIM Ambiguity Database [3].

2.3. Labeling of Ambiguity

Ambiguities occur at many different levels: dialogue, turn, and utterance level [10, pp. 2-3]. Ambiguities in the corpus (see Table 1, 2) were manually labeled according to the ambiguity labeling technique [10, pp. 1-7]. Each labeled ambiguity was, in turn, augmented with several attributes so as to make them easily manipulable [5, p. 2] . For example, in [3, p. 7], one utterance in which the fragment "yes" is ambiguous is represented as:

```
(ambiguity42-5.2 ((object "yes")
                  (status fatal)
                  (IFT (agreement discourse-indication information-offer))))
```

the numbers indicate its position within the corpus and its ambiguity category respectively. "Object" represents the actual fragment where the ambiguity arises, and "status" indicates the degree of ambiguousness. In this case, the value "fatal" means this fragment would be ambiguous even to a native speaker, IFT (illocutionary force type) represents the speaker's intention in saying the ambiguous fragment "yes." The same technique was used to label Japanese, English, and French corpus.

2.4. Distribution of Ambiguity

The following tables are presented here to give the readers an idea of the corpus used to build the database and the distribution of the ambiguities found in the corpus.

Table 1: Corpus used to build MIDDIM Ambiguity Database (Transcribed Dialogues) copied from [3, p. 5]

corpus name	language	no. of utterances	no. of words	no. of pause units
2a	Japanese	23	163	81
2b	Japanese	58	306	155
5a	English	21	328	69
5b	English	30	586	119
total		132	1383	424

Table 2: Corpus used to build MIDDIM Ambiguity Database (Written Texts) copied from [3, p. 6]

corpus name	language	type	no. of sentences	no. of words
CB5	English	abstract	3	86
CB6	English	introduction	13	443
ML7	French	abstract	9	162
CB8	French	introduction	11	452
GF9	French	transparency	12	50
GF10	French	transparency	10	59
total			58	1252

Table 3. Ambiguity Distribution
copied from [3, p. 17]

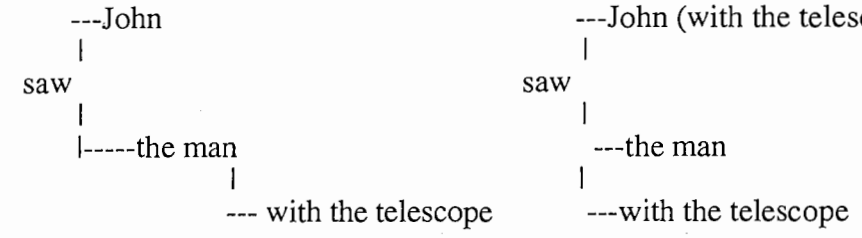
ambiguity class	Japanese	English	French	total (%)
categorical	1	2	0	3 (0.9)
structural	101	97	30	228 (75.7)
semantic	5	11	5	21 (6.9)
logico-semantic	0	2	3	5 (1.6)
discourse	19	20	5	44 (14.6)

Further analysis has been performed on four more dialogue files comprising a total of 391 turns. Interested readers are referred to [10].

3. INTERACTIVE DISAMBIGUATION STRATEGIES AND AN INTERACTIVE DISAMBIGUATION MODULE

As stated in the previous section, ambiguities should be resolved by guessing (in the case of humans) and automatically (in case of machines) as much as possible. When automatic disambiguation is not possible, the following strategies are conceivable [MIDDIM Minutes, 2/14/1994].

(1) Taking our "woman in the park" example, a disambiguation method could simply ask the question:
"Who is in the park ?"
(2) A second disambiguation method would present the user with a menu such as:
(a) Fred is in the park
(b) The woman is in the park
(c) Both Fred and the woman are in the park
(3) A third disambiguation method would produce an editable outline structure of the sentence, such as,



and the user could change this by moving the elements around.

(4) Still a fourth disambiguation method would be to animate the collection of all possible lexical trees of the sentence, and get the user to select the correct one.

The second disambiguation method, in fact, has been incorporated into an interactive disambiguation module [9]. A French version of the interactive disambiguation module was developed as a part of the Large Internationalization of Documents through Interaction with the Authors (LIDIA-1.0) mockup [6] of a Dialogue Based Machine Translation (DBMT) system [15]. An English version was created with the collaboration between ATR-ITL and GETA, and a Japanese version was suggested but has not been implemented.

In this module, the ambiguity resolution scheme is based on a pattern matching mechanism [6, p. 23]: that is, when the Ariane analyzer [13] finds an ambiguity (e.g., attachment problem) it generates multiple trees (called solution, Si in Figure 2) and assigns them to one of the four meta ambiguity classes [6, pp. 25-34] that are associated with one or more sets of ambiguity patterns called beams; the trees, then, are matched against the patterns; once matching patterns are found, disambiguation dialogues with items for the user to choose from appear on the menu. Choosing an item will trigger another round of pattern matching which will, in turn, generate another dialogue. This process continues until all the ambiguities are resolved. This module was used to generate disambiguation dialogues in the second and third experiments explained in the next section.

4. EXPERIMENTS

4. 1. First Experiment

The first informal experiment was conducted to gain a better understanding of (1) users' behavior in multimodal interactions, and (2) a possible interactive multimodal simulation environment specifically customized for Interpreting Telephone situations [7, p.2]. The aim of the this experiment was to make a "qualitative observation" rather than a statistical statement of the interaction modalities on the EMMI [2] environment. It was designed specifically "to test what influence variations in the speech recognition ability and language proficiency of the interpreter, and domain knowledge of all participants, will have on the number of disambiguation questions asked by the (automatic) interpreter [7, p.1]," and to collect ambiguities in this multimodal interpreting situation. The main independent variables were (1) language proficiency (English, Japanese, and French), (2) domain knowledge (algorithms, data structure, giving directions etc.), and (3) auditory aptitude (speech recognition ability : e.g., very good for the real human actors (agent, client) and average or fair for the machine (interpreter)), and side factor considered was environment mastery (familiarity with EMMI environment). The dependent variable was the number of disambiguation questions asked.

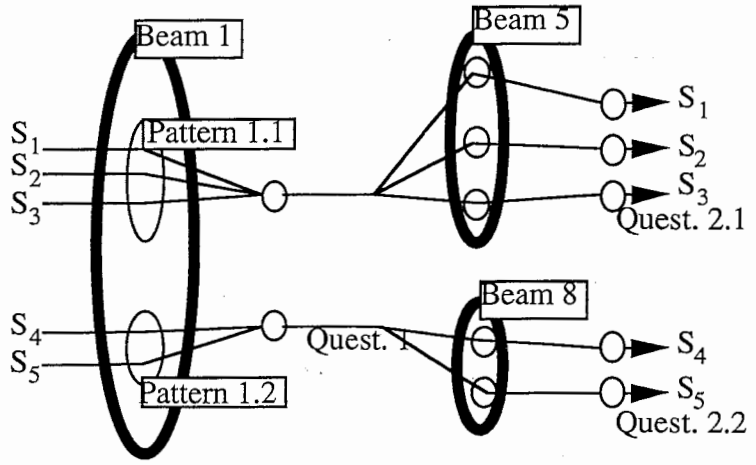


Figure 2. Overview of the disambiguation process

From the results of the experiment, the following observations were made: the language proficiency and the domain knowledge of the interlocutors will affect the number of disambiguation questions asked. However, it was evident from the results that people would adapt very quickly to a noisy auditory channel and a multimodal environment. If language proficiency is not perfect, any supplementary difficulty, such as problems with the environment, will lead to a considerable increase in disambiguation questions [from private correspondence between the first author and Boitet].

4.2. Second and Third Experiments

Two subsequent experiments were conducted to evaluate the understandability of the disambiguation dialogues generated by the aforementioned Interactive Disambiguation Module in comparison with dialogues generated in a more natural human-like manner [14], and to ascertain the effects of the different media (text vs. speech) on the subjects' ability to choose the correct answers [11].

Table 4. An example of human-like and machine-like disambiguation dialogues

human-like dialogue	machine generated dialogue
the cost of the taxi to the library	*the cost of the taxi to the library*
the cost of the taxi that is going to the library	the cost of the taxi towards the library
the cost of the taxi that will be paid by the library	the cost of the taxi for the library

The results of the second experiment showed that the subjects' ability to select the correct answers was not significantly degraded by the machine generated disambiguation dialogues. In other words, the machine generated dialogues were clear enough for the subjects to choose the correct answers.

The results of the third experiment showed that the subjects' ability to select the correct answers was not significantly affected by different media; i.e., disambiguation dialogues could be presented either in text or speech. Based on these results, the authors of [11] concluded that it is not crucial for those designing a disambiguation system to select the medium (speech vs. text) to generate the dialogues. The authors

further suggested adding a feature to the system by which users can select their medium of preference [11, p. 18].

5. SUMMARY AND CONCLUSION

In summary, the objective of this project was to investigate the optimum combinations of interactive disambiguation modalities (techniques) that would be adequate for written and spoken input in a multimedia context. Due to the part-time, long distance, and multi-culture basis of the project, we were not able to fully reach our objective within the time limit. However, our three years of collaborative research has enabled us to: perform three experiments, collect a sizable number of dialogues in three languages, observe real ambiguities arising in realistic situations which lead us to defining, classifying, and labeling ambiguities, and finally, produce a generic interactive disambiguation system. These three experiments were a learning experience through which we discovered the factors affecting the number of disambiguation questions asked, confirm the fact that machine generated disambiguation questions were clear enough for users, and measured the medium disparity between speech and text (no difference).

Acknowledgement

The author would like to express their sincere gratitude to Dr. Yamazaki of ATR-ITL and Professor Peccoud of GETA for their contributions in making this project possible. The author would especially like to thank Professor Boitet for his friendship, insightful advice, and consistent support throughout the entire project. Many thanks to Drs. Blanchon, Fafiotte, and Fais for spending countless hours running the experiments. Also special thanks to Ms. Tomokiyo without whom the MIDDIM-DB would not have existed; and finally, to Mr. Kurihara whose special programming skill always makes our research a much more tolerable endeavor.

REFERENCES

[1] Christian Boitet, Multimodal Interactive Disambiguation: first report on the MIDDIM project, ATR-ITL Technical Report TR-IT-0014, 8, 1993

[2] Kyung-ho Loken-Kim, Fumihiro Yato, Kazuhiko Kurihara, Laurel Fais, Ryo Furukawa, EMMI-ATR Environment for Multi-Modal Interactions, ATR-ITL Technical Report TR-IT-0018, 9, 1993

[3] Mutsuko Tomokiyo and Kyung-ho Loken-Kim, Ambiguity Analysis and MIDDIM-DB, ATR-ITL Technical Report TR-IT-0064, 8, 1994

[4] Geroges Fafiotte and Christian Boitet, Specification of first EMMI Experiments for the MIDDIM project in the context of Interpreting Telecommunications, ATR-ITL Technical Report TR-IT-0070, 9, 1994

[5] Christian Boitet, On the design of MIDDIM-DB, a data base of ambiguities and disambiguation methods, ATR-ITL Technical Report TR-IT-0072, 9, 1994

[6] Herve Blanchon, Pattern-based approach to interactive disambiguation: first definition and experimentation, ATR-ITL Technical Report TR-IT-0073, 9, 1994

[7] Geroges Fafiotte and Christian Boitet, Report on first EMMI Experiments for the MIDDIM project in the context of Interpreting Telecommunications, ATR-ITL Technical Report TR-IT-0074, 9, 1994

[8] Christian Boitet and Mutsuko Tomokiyo, Towards ambiguity labeling for the study of interactive disambiguation methods, ATR-ITL Technical Report TR-IT-0112, 4, 1995

[9] Herve Blanchon, An Interactive Disambiguation Module for English Input: an Engine and the Associated Lingware, ATR-ITL Technical Report TR-IT-0129, 8, 1995

[10] Mutsuko Tomokiyo, Ambiguity labeling for disambiguation experiments, ATR-ITL Technical Report TR-IT-0157, 3, 1996
[11] Herve Blanchon and Laurel Fais, A Second Experiment on the Understandability of Interactive Disambiguation Dialogues, ATR-ITL Technical Report TR-IT-00167,4, 1996
[12] Herve Blanchon, Kyung-Ho Loken-Kim, Tsuyoshi Morimoto, An Interactive Disambiguation Module for English Natural Language Utterances, NLPRS'95, pp. 550-555
[13] Christian Boitet, Software and lingware engineering in recent (1980-90) classical MT: Ariane-G5 and BV/areo/F-E/ Proc. ROCLing (tutorials), 8, 1990 in [page 23 in 6]
[14] Herve Blanchon and Laurel Fais, How to Ask Users About What They Mean: Two Experiments, ATR-ITL Technical Report TR-IT-00130
[15] Herve Blanchon, Perspectives of DBMT for monolingual authors on the basis of LIDIA-1, an implemented mock-up, Proc. Coling-94, 8, 1994, vol. 1/2, pp. 115-119

APPENDIX (MIDDIM'96 Workshop)

SESSIONS

1. Disambiguation: automatic or interactive, or both? Mon, 9:15—10:30

Chairperson: A. KUREMATSU, Animator: H. BLANCHON

A Chronicle of ATR-GETA multimodal Interactive Disambiguation Research Collaboration

Kyung-Ho LOKEN-KIM (ATR-ITL)

Abstract

A three year cooperative research project between ATR-ITL and GETA was conducted to investigate interactive disambiguation techniques of written and spoken input in a multimodal context. This paper first explains the scope of the project, second, summaries the joint research efforts made during the last three years in chronological order, and third, reports the outcome of the project.

Importance and Complementarity of Automatic and Interactive Disambiguation

Christian BOITET (GETA, CLIPS)

Abstract

Interactive disambiguation technology must be developed in the context of research towards practical Interpreting Telecommunications systems as well as high-quality multitarget text translation systems. In the case of speech translation, this is because the state of the art in the foreseeable future is such that a black box approach to spoken language analysis (speech recognition plus linguistic parsing) is likely to give a correct output for no more than 50 to 60% of the utterances ("Viterbi consistency"), while users would presumably require an overall success rate of at least 90% to be able to use such systems at all.

However, the same spoken language analyzers may be able to produce sets of outputs containing the correct one in about 90% of the cases ("structural consistency"). In the remaining cases, the system would be unable to analyze the input, or no output would be correct. Interactive disambiguation by the users of the interpretation or translation systems is then seen as a practical way to reach the necessary success rate.

In our view, interactive disambiguation is not to be used to solve all ambiguities. On the contrary, as many ambiguities as possible should be reduced automatically. The remaining ones should be solved by interaction as far as practically possible. What is left would have to be reduced automatically again, by using preferences and defaults.

Our stand is simply that, given the best automatic methods currently available, which use syntactic and semantic restrictions, limitations of lexicon and word senses by the generic task at hand, as well as prosodic and pragmatic cues, too many ambiguities will remain after automatic analysis, and the "best" result will not be the correct one in too many cases.

On the spot interventions expected (at least) from:

J.-Ph. GUILBAUD, T. MORIMOTO, K. MURAKI, M. TOMOKIYO, E. WEHRLI

2. Lexical disambiguation Mon, 10:45—12:00

Chairperson: E. BLANC, Animator: K.-H. LOKEN-KIM

Disambiguating Proper Names in Text

Yael RAVIN (IBM Research)

Abstract

Names occuring in text require special processing by most text processing applications: they cannot be translated, checked for spelling or interchanged with synonyms in the same way common words may be. In response to this need, we (at IBM's T.J. Watson Research Center in New York) developed Nominator - a module to identify and extract names from running text. Nominator consists of a fully automatic set of heuristics to locate names in text, determine what type of entity they refer to -- such as person, place or organization -- and group together all the variant names that refer to the same entity. These heuristics are based on typographical and lexical clues, with no parsing.

Just as ambiguous words have more than one sense, ambiguous names have more than one referent ("New York" refers to either the city or the state; "Mr. Reagan", to either the former president or his adviser). Just as ambiguous words can often be disambiguated by their context ("bank" is found either in the presence of other financial terms or in the presence of other geographical terms, for example), names can also be disambiguated by the presence of other names in their context: "IRA" will co-occur with "The Irish Republican Army" in one document and with "Individual Retirement Account" in another.

In the talk, I will discuss the kinds of reference ambiguities names exhibit and the heuristics we have devised to disambiguate them. I will also discuss how we are incorporating name disambiguation in our information retrieval system, both into the process of indexing documents and into the analysis of users' queries.

Interactive Lexical Priming for Disambiguation

William B. DOLAN & Stephen D. RICHARDSON (Microsoft Research)

Abstract

This paper describes an approach to word sense disambiguation which relies on user feedback to semantically prime a large lexical network when insufficient information is available to permit automated disambiguation.

Our work is implemented as a part of the Microsoft Natural Language Processing system, an effort aimed at producing an engine capable of mapping between arbitrary textual strings and structured, disambiguated semantic representations. Although our ultimate goal is fully-automated word sense disambiguation, there are clearly situations where this goal is not possible. For instance, short utterances with no associated semantic context often pose insoluble disambiguation problems. The task for our system is made especially difficult by our emphasis on broad coverage natural language processing: our English lexicon, derived from the online versions of the Longman Dictionary of Contemporary English (LDOCE) and the American Heritage Dictionary, 3rd edition (AHD3), encompasses more than 160,000 words and 240,000 word senses. Many words are highly polysemous; words with 50 or more senses are not uncommon

Consider the word "line", a notoriously ambiguous word in English which is often used as a test case for automated sense disambiguation. In our system, the combination of AHD3 and LDOCE yield a staggering 127 senses for the word "line", almost 120 of them noun senses. Presented with a bare sentence like "They put a line in", our system's automated sense-disambiguation component can do virtually nothing to identify the appropriate sense of "line" beyond ruling out the syntactically impossible verb senses. "Line" in this case could be plausibly resolved to almost any of the noun senses in our dictionary, including a row of objects, a piece of string, a mark on a page, a military formation, or a railway. While many of these senses share a metaphorical core of meaning, it is our assumption that extremely fine-grained sense distinctions may need to be made for tasks such as machine translation or information retrieval.

One way of handling this kind of situation might be to ask users directly for feedback about which sense is the appropriate one. For words like "line", however, the sheer number of senses would make this a frustrating task; at a minimum, work would have to be done to group these senses into semantically related homograph sets, with the user then being prompted to select from this reduced set. Arriving at a more precise sense assignment would involve forcing the user to descend through a hierarchy of senses. From our perspective, this sort of multi-step interaction with the user is

unacceptably burdensome. While acknowledging that interaction with the user is sometimes essential for adequate disambiguation, we would still like to keep these interactions as simple and infrequent as possible.

The solution that we have adopted within our system avoids the need to expose definitions of any kind to the user. Instead of requesting a technical decision about which sense or set of senses best corresponds to their intended meaning for a word, we simply ask the user to supply a word which describes or is closely related to the current discourse domain. This supplied word is used to "seed", or prime, MindNet, the huge lexical network which plays a crucial role in our system's sense disambiguation component.

MindNet is the product of automatically parsing and semantically processing the definitions and example sentences from both LDOCE and AHD3. Currently, this lexical knowledge base contains more than 300,000 instances of approximately 30 different types of semantic relations linking word senses. Paths connecting word senses are assigned weights based on a number of information-theoretic principles; these weights have been found to conform closely to human intuitions about relatedness.

This network and the tools our system provides for navigating through the semantic space it defines are proving invaluable for our system's automated sense disambiguation component, where it is used to model semantic priming. As undisambiguated content words are encountered during the processing of a coherent text, specific semantic regions of the network topology become activated, priming particular senses of polysemous words and influencing the behavior of the disambiguation component.

In the case of a sentence like "They put a line in", however, this crucial context is lacking. Failing in its attempt to choose a unique sense (or at least a sharply constrained set of possible senses) for "line", the system now asks the user to input one or more words closely related to current topic. Frequently, even a single word is enough to influence the state of the network in such a way that the appropriate sense(s) of the ambiguous word emerge from what would otherwise be chaos. Consider, for instance, how the 129 choices for "line" are restricted when MindNet is primed first with "army", then with "draw", and finally with "fish".

Word supplied by user: "army"

n, 29d, AHD3: The regular forces of any army or navy, in contrast to staff and support personnel.

n, 17a, LDOCE: (In the British army) the regular foot soldiers of the army (not the Guards).

n, 17b, LDOCE: (In the US army) all the regular fighting forces.

Word supplied by user: "draw"

n,3b, AHD3: A degree or circle of longitude or latitude drawn on a globe.

n,2, LDOCE: A thin mark with length but no width, which can be drawn on a surface.

n, 4c, AHD3: A contour or an outline.

Word supplied by user: "fish"

n 4, LDOCE: a cord with a hook at the end, usually fastened on a rod, used for fishing.

In our talk we will present a demonstration of our system's interactive disambiguation capabilities as well as MindNet.

On the spot interventions expected (at least) from:

J. CHAUCHÉ, M. DYMETMAN, S. EUBANK, N. NÉDEAU

3. Interactive Disambiguation in MT

Mon, 17:00—18:15

Chairperson: T. MORIMOTO, Animator: W. DOLAN

Transparency Encourages Users getting into Systems: Interactive Disambiguation in TWP

Kazunori MURAKI, Kiyoshi YAMABANA, Shinichiro KAMEI, Shinichi DOI (NEC labs)

Abstract

We focus on the role of interactivity in an interactive translation method, which enable user easily to get into system. We show that simple and familiar operations for inputting Japanese characters by utilizing Kana-Kanji conversion method, i.e. target word/phrase selection and specification of conversion scope, are powerful enough to resolve lexical, syntactic, and semantic ambiguities that were difficult for machines to solve. Since user has only to choose appropriate target language equivalent and proper meaningful phrase to translate, the burden of interactive ambiguity resolution is greatly reduced.

The availability of interactive disambiguation, more generally speaking, interactive systems can not exist without system transparency and direct manipulative feeling. Users in front of fully automatic MT system have experienced their heavy job of source text preparation including pre-edition and word registration ended in vain. The fact that pre-edition nor word registration never promise the happy translation, lets users to give up controlling translation quality by text preparation. They can not draw in mind an image how to manipulate MT as they want to have better results.

Our TWP(Translation Word Processor) way of interactivity secures transparency of system and direct manipulability to produce the result as users expect. We present the basic idea and describe the significance by referring to an English writing support capability of TWP.

Dialogue-Based MT for Monolinguals and future self-explaining documents

Christian BOITET (GETA, CLIPS)

Abstract

We are investigating DBMT (Dialogue-Based MT), a new architecture for MT systems enabling monolingual users to generate high-quality translations into several languages. The quality should be good enough to require no postedition or a minimal postedition in the case of very important documents. Fully automatic MT systems can achieve this goal only for very specific kinds of texts, such as weather bulletins.

In our approach, the document is first interactively "cleaned" and "tagged" (for special terms or idioms, utterance styles, etc.). The units of translation are then sent to a state-of-the art all-paths analyzer written in Ariane-G5 and running on a server, which returns a structure factorizing all ambiguities which cannot be reliably solved with the knowledge available. A question mark appears next to each ambiguous unit of translation. The user clicks on it to activate a disambiguation dialogue which doesn't suppose any particular expertise in linguistics, computer science, or any of the target languages.

Our LIDIA-1.2 mockup demonstates the concept on a HyperCard stack showing typical kinds of ambiguities in French. The target languages are Russian, German and English. The MT server resides on an IBM-9221 minicomputer which is accessed in the background by using a usual e-mail facility.

A reverse translation facility has been built into LIDIA-1.2 to study how to give the user some feeling of the target texts. Preliminary experiments have led us to the even more interesting concept of "self-explaining document": with the DBMT approach, it should be possible to produce all versions of a documents in an enriched format which would on demand show the ambiguous parts and the way they should be understood.

On the spot interventions expected (at least) from:

F. BOND, K.-S. CHOI, S. IKEHARA, M. STEDE, E. WEHRLI

4. Disambiguation at text & discourse level Mon, 18:30—19:45

Chairperson: J. CHAUCHÉ, Animator: M. STEDE

Local Cohesion as Contextual Constraints in Spoken Dialogues

Naoto KATOH (ATR-ITL)

Abstract

Discourse structure in task-oriented dialogues has two types of cohesion: global cohesion and local cohesion. Global cohesion is a top-down structured context and is based on a hierarchy of topics. On the other hand, local cohesion is a bottom-up structured context and a coherence relation between utterances, such as question-response or response- confirmation. This paper presents a method for recognizing local cohesion between utterances. We can automatically acquire discourse knowledge from an annotated corpus with local cohesion. In this paper we focus on speech act type-based local cohesion. The presented method consists of two steps 1) identifying the speech act expressions in an utterance and 2) calculating the plausibility of local cohesion between the speech act expressions by using the dialogue corpus.

Ambiguities in analysis by analogy

Yves LEPAGE (ATR-ITL)

Abstract

Analysis by analogy is a technique which bases on a possible computational definition of analogy. Its realisation requires the use of a tree-bank, and it is a kind of equation solving on a triple of trees. Results obtained in first experiments show that the same solution may be computed using different triples of trees. Building a better, more compact, tree-bank, which would be representative of the same amount of information, is tantamount to minimising this kind of non-determinism. Also, analysis by analogy generally delivers different solutions. We shall study whether these different solutions correspond to some relevant linguistic facts, whether they reflect ambiguity in the tree-bank (inconsistency of coding), or whether they come from the algo- rithm used to solve analogy equations.

Interactive Disambiguation in Multilevel Parallel Texts Alignment

TANG Enya Kong(UTMK, USM, Penang)

Abstract

In this paper, we describe a framework for aligning different type of translation units exist in the parallel texts semi automatically, namely aligning with human help by mean of interactive disambiguation. The type of translation units treated here not only consist of basic text units like sentence, phrase and word, it also takes syntactic dependency structure into consideration. Sentences are aligned based on a simple statistical model of character lengths guided by the paragraph structures of the parallel texts. Whereas those components internal to the aligned sentences, namely phrases and words together with their dependency relations are aligned based on :

- the measure of similarity between the distribution of words (phrases) in the aligned sentences of the parallel texts.
- the concordances of special words and cognates in the aligned sentences of the parallel texts.
- the bilingual (both general and domain specific) dictionary.
- the POS tag assigned to each word in the parallel texts.

All the above mentioned alignments can be kept effectively in terms of the synchronous Structured String-Tree Correspondence (SSTC) [Boitet&Zaharin 88].

One way of handling this kind of situation might be to ask users directly for feedback about which sense is the appropriate one. For words like "line", however, the sheer number of senses would make this a frustrating task; at a minimum, work would have to be done to group these senses into

semantically related homograph sets, with the user then being prompted to select from this reduced set. Arriving at a more precise sense assignment would involve forcing the user to descend through a hierarchy of senses. From our perspective, this sort of multi-step interaction with the user is unacceptably burdensome. While acknowledging that interaction with the user is sometimes essential for adequate disambiguation, we would still like to keep these interactions as simple and infrequent as possible.

On the spot interventions expected (at least) from:

S. IKEHARA, M. SELIGMAN, M. TOMOKIYO

5. Ambiguity definition, classification, labelling Mon, 9:15—10:30

Chairperson: N. NÉDEAU, Animator: K. MURAKI

On the formal definition of ambiguities and associated concepts and an associated ambiguity-labelling scheme

Christian BOITET (GETA, CLIPS) & Mutsuko TOMOKIYO (ATR-ITL & GETA, CLIPS)

Wanting to build a data base of ambiguities in the context of the MIDDIM project, we have been struck by the realization that, although the notion of ambiguity is very familiar, we did not find any single fromal definition precise enough for our purposes. In this presentation, we make more precise several notions, such as ambiguous representation, ambiguity, ambiguity kernel, ambiguity type, etc.

First, we speak of *ambiguity of segmentation into turns* or *into paragraphs* if in a dialogue (resp. a text) if it can be segmented in at least two different ways into turns (resp. paragraphs). Similarly, we speak of *ambiguity of segmentation into utterances* in a turn (resp. a paragraph) if it can be segmented in at least two different ways into utterances.

Second, we speak of *utterance-level ambiguitiy* in an utterance if it can be analyzed in at least two different ways, whereby the analysis is performed in view of translation into one or several languages in the context of a certain generic task. The question is more complex, as ambiguities of such kind may be cause by the presence of several competing interpretations, which are encoded in various ways by analyzers, as well as by the absence of some information necessary in the considered context, such as the definiteness of nominal groups in translating from Russian, Chinese or Japanese into French or English.

We then go on to define the support of an ambiguity as the minimal substring giving rise to an ambiguity of multiplicity or underdefiniteness *in the context of the whole utterance.*This enables us to define ambiguities as objects, and to propose a scheme for ambiguity labelling, where each instance of ambiguity is characterized by its scope, its linguistic type and subtypes, and various characteristics such as its importance, its status, etc.

Experiments in ambiguity labelling of dialogue transcriptions and texts in English, Japanese and French

Mutsuko TOMOKIYO (ATR-ITL & GETA, CLIPS) & Monique AXTMEYER (GETA, CLIPS)

We explain in more detail the labelling scheme we have used, as well as the linguistic characteristis of ambiguities we have defined and used in our experiments. We will present and discuss various examples. We try to show how it is possible, although not always straightforward, to perform that kind of labelling with the goal to label only the ambiguities which we estimate state of the art , robust automatic analyzers would produce.

On the spot interventions expected (at least) from:

H. BLANCHON, Y. LEPAGE, N. NÉDEAU, M. SELIGMAN

6. Solving ambiguities: methodology & tools Tue, 10:45—12:00

Chairperson: L. ZAYSSER, Animator: E. WEHRLI

Merging statistical and rule-based methods using decision trees

Stephen EUBANK (ATR-ITL)

Abstract

Rules generated by an expert can easily be tested for relevance and usefulness using decision tree (CART) models. The resulting synthesis is more powerful than either automatically-generated or expert-generated rules alone. The benefits of this approach are: systematic, unbiased selection of rules; efficient use of data; and automatic derivation of complex combinations of features. For example, we have constructed a part-of-speech tagger in a two stage process: first, a linguist proposes a large (currently around 2000) set of "features" of the context which he considers relevant for tagging; second, a decision tree is allowed to pick and choose from among the proposed features. The resulting tagger gives state-of-the-art performance on a standard tag set. This procedure can easily be extended to more detailed tag sets which are more appropriate for automatic disambiguation.

A Customizable Interactive Disambiguation Methodology and Two Implementations to Disambiguate French and English Input

Hervé BLANCHON (GETA, CLIPS)

Abstract

As natural language is highly ambiguous even in restricted domains, interactive disambiguation is seen as a necessity to achieve more robust and user-friendly interactive systems, face to face translation, systems and Dialogue-Based Machine Translation systems. We have proposed a methodology which distinguish to part in a disambiguation module: an engine (language and application-independent) and a lingware (language and application -dependent). The engine should, thus, be reused in the design of any disambiguation module. This paper present the current state of or work, that is: an engine that has been used to design two interactive disambiguation modules, one for French and one for English.

On the spot interventions expected (at least) from:

B. DOLAN, J.-Ph. GUILBAUD, K.-H. LOKEN-KIM, M. STEDE

- **— Excursion & banquet —** Tue, 13:30—

Visit of "la Correrie" (Chartreuse monastery), the Voiron Caves (where the Chartreuse liquor is produced) and the St Hugues chapel on the Chartreuse plateau, well-known for its beautiful decoration. Banquet in the mountains, close to the hotel.

7. Software design for disambiguation Wed, 9:15—10:30

Chairperson: Y. RAVIN, Animator: M. SELIGMAN

Representation of morpho-syntaxical ambiguities

Laurence ZAYSSER (GSI-Erli, Paris)

Abstract

The grammars for morphological disambiguation introduce a change of axis in the course of the analysis: with them, one goes from the paradigmatic to the syntagmatic, although the latter is reduced to its simplest expression. This is not without consequence on the representation of the data.

Software engineering in the LIDIA project: distribution of interactive disambiguation and other components between various processes and machines

Mathieu LAFOURCADE (GETA, CLIPS & UTMK, USM)

Abstract

In the framework of the LIDIA project - which aims at designing and implementing a Translation Workstation for the Author - we describe how the interactive disambiguation can be viewed as a

collection of concurrent processes. These processes can be either locally distributed - on the same computer - or widely distributed in the Internet. We propose and compare two approaches for the communication, one by e-mail (NMTP) and the other by the protocol used in the World Wide Web (HTTP). Although really robust, the e-mail based communication is applicable only to tasks that can be slightly deferred. The HTTP scheme can better be used in contexts where time response is more an issue. We show how such processes can be seen as services usable by several clients irrespective of their actual applications.

On the spot interventions expected (at least) from:

H. BLANCHON, Ch. BOITET, Y. RAVIN, M. STEDE

8. Human factors & multimedia aspects of ID Wed, 10:45–12:00

Chairperson: S. IKEHARA, Animator: K.-S. CHOI

An analysis of the first EMMI-based Experiments on Interactive Disambiguation in the context of Automated Interpreting Telecommunications

Georges FAFIOTTE and Christian BOITET (GETA, CLIPS)

During our research stay at ATR, we experimented for the first time with EMMI, ATR multimodal platform, in a situation involving a client (C) and an agent (A) speaking two different languages and not knowing the other, and a human interpreter (B) playing the role of an ideal Speech Tanslation sustem (Wizard of Oz situation).

In the first experiment, the goal of A was to explain a complex algorithm and the associated program to C, and B had no knowledge of the subject matter. In the second experiment, A explained to C how to get from Lyon airport to our campus, whereby B did not know the precise topography of the region but had obviously a general competence about airports, shutlle buses, tramways, maps, etc.

In total, these two experiments were broken down into seven phases. We tried to let certain factors vary, such as the linguistic proficiency of B in one of the two languages, and the domain knowledge of B, and to a lesser extent C. However, the linguistic proficiency of B must in all cases be quite high, so that B can interprete in the first place.

We found it quite difficult to artificially lower the linguistic proficiency of B by using acoustical distortion, as humans adapt to it very rapidly. The influence of the "environment mastering" factor was more visible, as humans don't adapt so quickly to a complex set of windows, icons, buttons, images, forms, cameras, buttons, etc. Finally, the domain knowledge (or rather lack of knowledge) of B crucially increased his/her tendency to ask clarifying questions. We could also verify the quite natural hypothesis according to which B will suddenly start to ask many questions if his/her cognitive load exceeds a certain limit, due to a combination of lack in domain knowledge, environment mastery, and (to a lesser extent) linguistic proficiency.

How to Ask Users About What They Mean: Two Experiments and Results

Hervé BLANCHON (GETA, CLIPS) & Laurel FAIS (ATR-ITL)

Abstract

We describe two experiments we have conducted at ATR using "naive" subjects and the Wizard of Oz technique, in two different contexts. English was always the source language.

The subjects were shown ambiguous sentences from our data base, which contains, for each sentence, a set of analysis trees. Disambiguation dialogues were automatically generated using a specialization of our generic disambiguation engine to English and that kind of trees.

The reactions of users were collected and evaluated with respect to the types of ambiguities and of disambiguation questions. Although some ambiguities still seem difficult to understand by naive users, almost all sentences were correctly disambiguated, with favourable reactions from the users.

On the spot interventions expected (at least) from:

J. CAELEN, B. DOLAN, K.-H. LOKEN-KIM

9. Disambiguation cues for speech translation Wed, 17:00—18:15

Chairperson: M. TOMOKIYO, ~~Animator: Y. LEPAGE~~

Ambiguities in Multimodal Machine-interpreted Dialogues

Laurel FAIS ~~(ATR-ITL)~~

Abstract

In this paper we will examine the occurrence of ambiguities in interpreted, task-oriented dialogues. Transcriptions from human-interpreted and simulated machine-interpreted conversations were examined for ambiguities. Using the categories described in Blanchon (1994), we describe the nature and frequency of occurrence of six categories of ambiguities as found in these dialogues. In the machine-interpreted conversations from which some of these transcriptions were made, subjects were asked by the "machine" to repeat complex or lengthy utterances. These utterances sometimes contained ambiguities. The nature of the repairs subjects made in this situation is discussed with respect to the spontaneous disambiguation of ambiguous utterances.

Disambiguation in VerbMobil (to be confirmed)

Manfred STEDE (TU, Berlin)

Abstract: *forthcoming*

Prosodic cues for speech disambiguation

Geneviève CAELEN-HAUMONT (GEOD, CLIPS)

Abstract: *forthcoming*

On the spot interventions expected (at least) from:

J. CAELEN, N. KATOH, S. EUBANK, E. KELLER, K.-H. LOKEN-KIM

10. Interactive architectures for speech translation Wed, 18:30—19:30

Chairperson: TANG E.-K., Animator: E. KELLER

Interactive MT and Speech Translation on the Internet

Mark SELIGMAN (Red Pepper & Intellicorp)

Abstract

I'll discuss various practical issues facing real-time MT communication over the Internet. One possible translation vehicle is Internet Relay Chat, or IRC, which relays text from user to user. Both one-to-one and many-to-many conversations are presently common. I'll propose that interactive MT can serve as a front end for such communication: a user would first interactively facilitate or debug translations in the source language, and when feedback had assured a translation of sufficient quality, would pass the translated text string to chat for transmission.

I hope to organize experiments with such a setup -- stressing questions of usability -- and will present (and solicit advice on) a range of ideas and problems. A second stage of experimentation would add interactive speech recognition as an additional front end (and perhaps speech synthesis as a back end). Both isolated-word and connected speech recognition will be considered: in either case, interactive validation seems crucial at the present stage of the technology.

The central proposal will be that while unsupervised SR and MT are still at the stage of basic research, supervised SR and MT may well be nearing the point of practicality, so that a good interface between them might yield a "quick and dirty" speech translation system in the near term. At the same

time, the network explosion presents an unprecedented opportunity for relatively low-cost experimentation. Central questions: For specific communication tasks, how much interaction is needed for what degree of quality, and how much quality is enough?

Place and degree of Interactive Disambiguation in Future Practical Speech Translation Systems

Christian BOITET (GETA, CLIPS) & Kyung-Ho LOKEN-KIM (ATR-ITL)

Abstract

It has always been remarkably difficult to build really practical Machine Translation (MT) and Speech Processing (SP) systems. As Speech Translation (ST) combines the difficulties of both endeavours, it should come as no surprise that the current prototypes based on a black-box paradigm cannot be extended towards practical systems, although they are well-researched and brilliantly demonstrated.

Dramatic progress in both MT and SP technology is not being likely to be witnessed in the near future. Besides necessary but inherently limited improvements in the component technologies, the construction of practical ST systems will require better user-friendliness, achievable through the introduction of a human expert (interpreter), multimodal communication facilities between the expert and the speakers, and various control and feed-back facilities.

Because of the quality and coverage required of the Speech Recognition (SR) and Natural Language Analysis (NLA) components in realistic applications and their inherent difficulty, however, it will also be necessary to involve the end users (the speakers) in these processes, by encouraging them to control their own voice, and asking them to help through multimodal active and passive disambiguation.

It should also be recognized that a whole palette of automation levels is available in the context of speech translation. They range from machine aids for direct bilingual communication to the above mixtures of automatic and human translation through on-line helps for human translators. Perhaps the best way towards practical systems would be to incrementally build the corresponding tools and environments, beginning with the less ambitious ones.

On the spot interventions expected (at least) from:

J. CAELEN, E. KELLER, G. FAFIOTTE, A. KUREMATSU, T. MORIMOTO, K. MURAKI, N. NÉDEAU, E. WEHRLI

- **Conclusions & perspectives** Wed, 19:30—19:45

Ch. BOITET, T. MORIMOTO, K.-H. LOKEN-KIM, E. BLANC

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

INDEX

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