TR-I-0165

ABMT for Text and Dialogue:
a preliminary assessment of its potentials

Toon Witkam

August, 1990

Abstract

ABMT (Analogy-Based Machine Translation) is a relatively new MT paradigm. Its most advanced concretization today is the BKB (Bilingual Knowledge Bank), in which parsed parallel corpora of coherent texts form an example database. This corpus-based approach enables an ABMT system to handle the *uncompositionality* inherent to translation.

The report focusses on the feasibility of the BKB method to *dialogue translation* as opposed to text translation. It appears that the problems addressed by the BKB method are largely common for both. Also, the method has potential for inferencing as part of dialogue understanding.

Two steps for further ABMT exploration are proposed: the extension of a small BKB study model with Japanese, and stochastic simulation of BKB operation.

ATR Interpreting Telephony Research Laboratories

ABMT for Text and Dialogue:
a preliminary assessment of its potentials

Toon Witkam*

ATR Interpreting Telephony Research Laboratories
Sanpeidani, Inuidani, Seika-cho, Soraku-gun
Kyoto 619-02, Japan

Abstract: An overview is given of recent work in ABMT (Analogy-Based Machine Translation), a relatively new MT paradigm which relies on approximate matching of previous translation patterns. Its most advanced concretization presently is the BKB (Bilingual Knowledge Bank), in which parsed parallel corpora of coherent texts make up the example database. After prefabrication of this BKB, which involves not only parsing but also semi-automatic identification of productive bilingual 'translation units', translation will be entirely corpus-based, as the BKB incorporates grammars, dictionaries, world-knowledge and (human) translation skill. The latter enables an ABMT system to handle -much better than mainstream MT systems- the *uncompositionality* inherent to the translation problem.

The report focusses on the question whether ABMT, and in particular the BKB method, offers prospects to *dialogue translation.* The answer is positive. It is argued that the new paradigm is part of basic technology, that there is cross-fertilization between written-text MT and spoken-language MT, and that there is considerable common ground between the two. In addition, the BKB method appears to have something in promise for non-MT inferencing, as part of building an internal meaning representation for goal-directed dialogues.

The report advises ATR to give continued attention to ABMT, and to proceed with BKB exploration. Two small steps for the coming year are proposed: the extension of a small (20.000 words) BKB study model with Japanese, and stochastic simulation of BKB run-time operation.

Key words: ABMT, corpus-based translation, BKB, basic technology, semantic distance, pattern matching, uncompositionality

---

* Invited researcher from BSO/Research, P.O.Box 8348, NL-3503 RH Utrecht, Netherlands

Table of Contents

1. Introduction

In this report, we discuss the role and prospects of ABMT as part of the basic technology for language and speech translation. ABMT (Analogy-Based Machine Translation) is a technique, remarkably different from conventional mainstream MT methods, which are largely rule-based (RBMT).

Analogy-based translation is closely related to translation-by-example, which has been introduced as an object of initial study at ATR[1] in 1989, and to which Nagao's group at Kyoto University has meanwhile demonstrated its affinity.

Views expressed in this report have been formed to a large extent on the basis of previous research experience in BSO, where considerable work on ABMT has been done during the past three years.
In addition, concrete proposals[2] for further exploration of ABMT are given, partly based upon a recent simulation model developed at BSO.

2. ABMT: new prospects for (textual) MT

This section contains the motivation for developing a new MT technique, ABMT. Analogy-Based Machine Translation is a technique intended for MT in general. Until now, BSO explored this technique primarily with translation of *written texts* in mind. The additional motivation for using ABMT in *dialogue* translation will be discussed in Section 3.

2.1. Why care about text MT?

Why should we care about textual MT in a research framework primarily aimed at telephone interpration?

Firstly, we think that, despite the differences (as highlighted by see Tsujii and Nagao, 1988[3]) the common grounds between written texts and spoken dialogues should not be forgotten. The *cohesion* brought about by a textual structure applies to dialogues as well as to texts. In section 3 we will give some technical arguments to this extent.

Secondly, we oppose slight suggestions (e.g. as read between the lines of Somers 1989[4]) that only in *spoken* language translation there were scope for significant advances (3rd generation).

---

[1] see also: Kurematsu, "An Overview of ATR Basic Research into Telephone Interpretation", Keynote Speech ASTI'89 (=ATR Symposium on Basic Research for Telephone Interpretation, 11-12 December 1989, Kyoto)

[2] section 5 of this document

[3] "Dialogue Translation vs. Text Translation - Interpretation-Based Approach", COLING'88, pp. 688-693; see also section 3 of the current document

[4] Harold L. Somers, Contribution to Panel Discussion "Future direction of language processing for automatic telephone interpration", ASTI'89, pp. 8-3-2, 8-3-3

In our view, the language technology for *written* text (pushed by the competition among *wapuro* manufacturers, and pulled by the demands of the international product documentation market) will not stay behind. Speech technology will benefit from advances in language technology - and the other way around.
Certainly if we are concerned with **basic technology**, the separation between the two technologies should not be a very tight one.

A third consideration is that even at the application level, we are not so sure about sharp distinctions between texts and dialogues in the electronic information age. Also this issue will be elaborated in section 3.

The potential of ABMT typically lies in handling those problems that are very hard to catch by rule-based formalisms. A notable problem when translating into English is *NP definiteness*; another one is the use of the *future tense*. These phenomena apply to text as well as to dialogues. Yet they have been identified as two of the three remaining "very difficult problems" for ATR's experimental speech translation system[5].

2.2. Why a new paradigm?

Spending effort on a new paradigm can be justified by the following background situation in MT research:

- the semantic barrier to high-quality MT is still there!
- prominent MT researchers signal consensus about failure (Wilks, King, ...'89)[6] of traditional methods such as: semantic primitives, deep cases, formal logic;
- *"Nobody knows how to organize a large body of knowledge for MT"* (Nagao,'88)[7];
- the need for **new** knowledge sources is emphasized (Kay[8]; Tsujii, Carbonell[9]).

Also if we look at the wider field of NLP (Natural Language Processing), of which MT is a part, the need for breakthroughs, for new paradigms, is made clear (e.g. by Obermeier, '89)[10]. There are already several new directions visible: probabilistics (Shieber[11]), neural networks and machine learning. A nice example of a totally new way of thinking about MT is Tsujii's envisaged "MT without source texts" (Tsujii[12]).

The growing dissatisfaction with the mainstream (rule-based, transfer-based) approach to textual MT goes together with a changing attitude about the role of linguistics, or at

---

5 Kurematsu,Iida,Morimoto,Shikano, "Language Processing in Connection with Speech Translation at ATR Interpreting Telephony Research Laboratories", SP Comm Journal 1990

6 IFTT'89, Oiso.

7 closing panel, COLING'88, Budapest.

8 closing panel, 1st MT Summit, Hakone '87.

9 IBM Europe Seminar on Computers and Translation, Garmisch-Partenkirchen, August '89.

10 Klaus Obermeier, "NLP and AI - the industry perspective", Horwood, '89.

11 probabilistics as the new, corpus-based trend in grammar; source: see footnote 4

12 IFTT'89, Oiso.

The growing dissatisfaction with the mainstream (rule-based, transfer-based) approach to textual MT goes together with a changing attitude about the role of linguistics, or at least computational linguistics, in general. Maybe COLING'88 will prove to be a historic turning point in this regard. In a panel called "Language Engineering: The Real Bottle Neck of NLP"[13] the sense of crisis and the need for a new course was presented by IBM's

HPSG Head-driven Phrase Structure Grammar
GPSG Generalized Phrase Structure Grammar
APSG Augmented Phrase Structure Grammar
CCG Combinatorial Categorial Grammar
FUG Functional Unification Grammar
UCG Unification Categorial Grammar
LFG Lexical/Functional Grammar
Co-representational Grammar
DCG Definite Clause Grammar
Relationally-Based Grammar
TAG Tree Adjoining Grammar
Transformational Grammar
Stratificational Grammar
Configurational Grammar
Neostructural Grammar
Phrasal Core Grammar
Equational Grammar
Functional Grammar
Systemic Grammar

Fig. 1. Proliferation of grammar formalisms.

Karen Jensen (" Why Computational Grammarians Can Be Skeptical About Existing Linguistic Theories"), by Roesner ("Why Implementors of Practical NLP Systems Can not Wait for Linguistic Theories"), by Tsujii ("Why I do not care grammar formalism") and by Tomita ("'Linguistic ' Sentences and 'Real' Sentences"). The latter highlighted the contrast between what linguists usually find interesting and what real-life texts look like[14]:

'Interesting, linguistic' problems:

John hit Mary.

The mouse the cat the dog chased ate died.

John persuaded Mary to expect that he believes that she likes an apple.

---

[13] organized and chaired by Nagao; see COLING'88 Proceedings, pp. 448-453

[14] the examples given here are from Tomita

'Uninteresting, real' problems:

> **This window contains an HP-UX shell (either a Bourne shell or C-shell, depending on the value of the SHELL environment variable; for details, see the "Concepts" section of the "Using Commands" chapter).**

In the same vein, American MT guru Martin Kay stated[15] on an earlier occasion that, despite the presence of a dozen or so grammar models in his Palo Alto vicinity, there still did not exist "a decent parser for English".
This situation, it appears, has not yet changed much. The list of grammar formalisms (fig. 1) continues to grow, but that does not really help us to pursue the following objectives:

- to break *the semantic MT barrier* (an already long-existing motivation);

- to meet the quality level of a *professional human translator* (a rather new motivation; usually, comparison with professional human translator's quality tends to be eschewed in MT[16], instead of taken as its "acid test");

- to design a method for (large-scale) *industrial manufacturing* of MT systems, not a method that lends itself primarily for thesis writing or laboratory experimenting; this point will be taken up again in Chapter 5 of this document.

Pursuing these objectives coincides with the engineering approach to MT (as advocated since COLING'88[17]), and also with the new 'trend'[18] toward interactive editing instead of post-editing.

In order to meet the above objectives, a radically new paradigm is necessary.

2.3. Worldwide history of ABMT research

In hindsight, the history of ABMT (to which we also reckon EBMT and MBT) begins in 1984, with the publication of Nagao's seminal paper: "A Framework of a Mechanical Translation between Japanese and English by Analogy Principle"[19]. It included a number of key elements, reproduced here concisely:

- need for fundamentally different MT approach
  (mainstream MT being at its boundary)

- "language learner" paradigm
  (learning by example sentences, not by rules)

---

15 closing session 1st MT Summit, Hakone '87.

16 an indication for this is the lack of any work on human translation in the otherwise well-equipped libraries of some MT research centers

17 Tsujii at IBM Europe Seminar on Computers and Translation, Garmisch-Partenkirchen, August '89: "We need a software engineering approach to MT".

18 Tsujii, Carbonell (ibid.)

19 in: Artificial and Human Intelligence - Elithorn and Banerji, eds., 1984.

- paradigmatic variation
  (replacement of a word in context)

- machine learning
  (training of the machine by a human teacher)

- extensional representation of word meanings
  (defining word groups by sentential context)

- analogy principle
  (similarity between given input and example)

- translation unit:
  from a block of words to a block of words

- machine learning, knowledge augmentation
  (rely on primary data, i.e. data which is not very much processed, which is independent of changes in linguistic theory)

All these key elements can be found back in BSO's later work (see 2.4), although Nagao's paper remained unknown to BSO for many years. It even remained unnoticed by the international MT community at large. Nagao himself, despite his presence at most MT conferences, hardly mentioned his early off-mainstream speculation, neither in speech nor in writing[20], during the period 1984-1989.
In 1989, Nagao affirmed[21] that ABMT was a very interesting line of research along which quite some research work could be expected in the next few years. In 1989 and 1990, two papers from Nagao's laboratory at Kyoto university appeared, both titled "Memory-based Translation"[22] (MBT).

Meanwhile, Sumita and Tsutsumi at IBM's Tokyo Research Laboratory had published[23] "A Translation Aid System Using Flexible Text Retrieval Based on Syntax-Matching", describing the first modest but practical attempt to run MT on previous translation examples.

As known, this work was followed by a more principle-oriented publication from ATR in 1990: "Example-Based Approach in Machine Translation" (EBMT), of which Sumita is the first author[24].

---

20 There is no mentioning at all of the analogy principle in Nagao's 1986 book "Kikai hon'yaku wa doko made kano ka", nor in the preface to its 1989 English edition: "Machine Translation, How Far Can It Go?"

21 during the closing discussion of Oiso's IFTT'89 conference, in reaction to a brief presentation, by the present author, of BSO's BKB work

22 both papers authored by Satoshi Sato and Nagao; the 1990 paper was presented at COLING'90, Helsinki.

23 Sumita and Tsutsumi; IBM TRL Research Report TR87-1019, 5-12-'88

24 co-authored by H.Iida and H. Kohyama; published at the 3rd MT Conference, Austin, Texas, June 1990; also to appear in InfoJapan'90 at Tokyo, October 1990.

As to related research: a strong relation with ABMT (though not labelled as such) can be found in recent thinking at AT&T Bell Labs, considering a contribution from Church et al.[25] They argue in favor of:

- syntactic preprocessing of large corpora ('statistics alone is not enough');
- statistical assessment of word associations ('typical predicate-argument relations') as opposed the use of IS-A hierarchies;
- statistical assessment of collocations (instead of using purely syntactic methods).

All this is very supportive to BKB-style ABMT (see 2.4).
Furthermore, a recent thesis by Laffling[26] gives support to translation by analogy. It shows that existing bilingual dictionaries are inadequate, and that lexical transfer should instead be based on equivalences extracted from *(globally, not strictly) parallel* cohesive texts in the target language.
Less recently, the work by Sampson, also in Britain[27], can be considered as pioneering in rule-less parsing.

2.4. ABMT elaboration: BSO's BKB method

At BSO in the Netherlands, R&D on ABMT evolved from 1987. In that year, the explicit semantics (a priori hierarchy, inheritance) previously[28] exercised in BSO's Research Labs at Utrecht was replaced by implicit (analogue) semantics, characterized by an operationally well-defined *semantic distance* function, based on statistics of contextual overlap. It was embodied in BSO's 1988 English-French DLT prototype.

During the period 1988-1990, a particular concretization of ABMT in the form of the so-called BKB (Bilingual Knowledge Bank) structure was designed. This work was largely inspired by the need for a *manufacturing technique*, in view of future needs for a diversity of languages and domains in industrial MT. The particularity of the BKB design is that it combines linguistic and extra-linguistic knowledge sources into one structure.

These developments at BSO were part of the DLT (Distributed Language Translation) project for written-text MT, which involved some 50 person-years in the period 1985-1990. The semantic distance and the BKB are inventions by Dr. Victor Sadler, the team's senior semanticist. Patents for the semantic-distance method and the BKB design were first filed in the Netherlands early in 1989, and submitted in other countries (including Japan) in 1990. Extensive publication in book form occurred at the end of 1989[29].

---

[25] Kenneth Church, William Gale, Patrick Hanks and Donald Hindle, "Parsing, Word Associations and Typical Predicate-Argument relations", ASTI'89.

[26] Laffling: "Machine Disambiguation and Translation of Polysemous Nouns - a lexicon-driven model for text-semantic analysis and parallel text-dependent transfer in German-English translation of party political texts", University of Birmingham, April 1990; to be published.

[27] Geoffrey Sampson, University of Leeds (probabilistic models of analysis, simulated annealing etc.)

[28] period 1985-1986.

[29] Sadler, "Working with Analogical Semantics", FORIS 1989

Central in BSO's approach is the use of parallel[30] text corpora as an all-encompassing knowledge source. They serve as a large and very rich example database, to such extent that no separate dictionary is needed. At the same time, these corpora are not collections of isolated examples, but cohesive texts. After parsing into a simple basic representation of word-dependency trees and imposing a 2-dimensional pointer structure on it, the so-called BKB becomes sort of a "power engine" for ABMT:

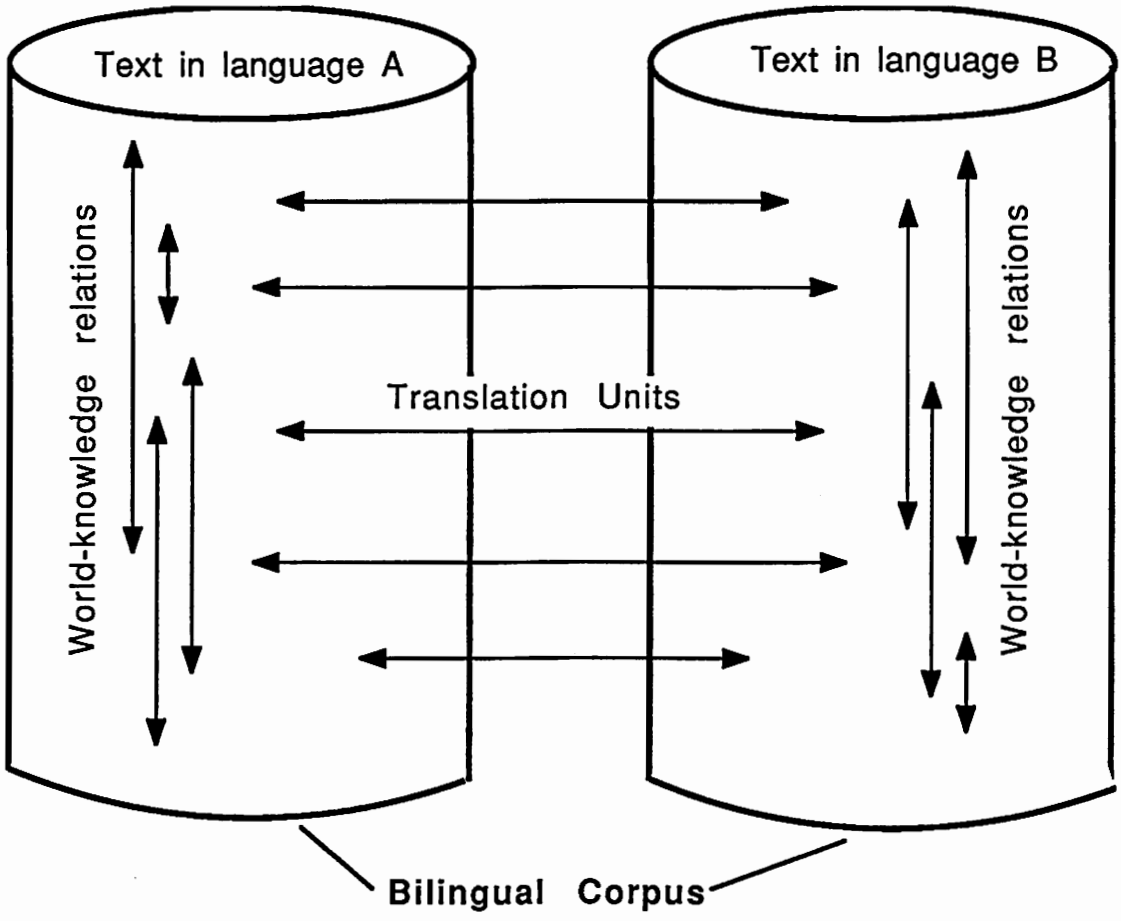

Fig. 2. BSO's translation and knowledge bank, usually referred to as BKB (Bilingual Knowledge Bank).

Whereas other approaches to ABMT (section 2.3 above; see also Table 1) feature the use of thesauri or suggest a hybrid mode, in which ABMT supplements RBMT, the design by BSO is more radical, and therefore totally different from 'mainstream' MT (most European and Japanese transfer-based and interlingual systems, viz. METAL, EUROTRA, ATLAS, SHALT, PIVOT...). In BSO's method, after prefabrication[31] of the BKB,

[30] i.e. in 2 two languages, the one being a reliable translation of the other

[31] BKB prefabrication does involve linguistic analysis, but also this is largely done without recourse to conventional rule-based or dictionary-based methods.

no linguistic analysis or decomposition, in the usual sense, takes place. The translation is exclusively done by pattern matching of previous translations.

MT as an information retrieval task rather than a linguistic problem is not an entirely new idea; a prominent MT researcher as Tsujii [32]plays with it, and the idea has been suggested once by an MT-layman in EEC circles[33]. To get a good appreciation of all this, let's draw an axis (fig. 3) with two extreme positions at its ends. At one side, we have the simple phrasebook-type systems of pairs or sets of corresponding sentences. These paper or pocket-electronics systems are sometimes arranged with the help of multilingual number codes, and might have fill-in-the-blanks provisions for proper names or nouns unknown to the system. They are generally considered as naive and totally unscientific by the research community.

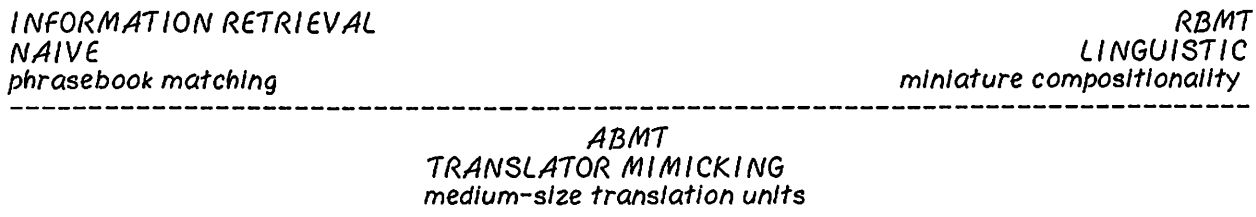

Fig. 3. Axis of different paradigms and technologies to approach MT.

At the other end of the scale we have the type of MT system developed by formal linguists: RBMT, based on a grammar and dictionary model out of the theoretical linguist's laboratory, nearly mathematics, with miniature compositionality as its primary objective[34].

The contrast is between matching whole sentences (or even entire paragraphs or documents[35]) and matching only the smallest elements (words or morphemes, as in the standard RBMT approach). In case of the former, there is little or no compostion problem, but the productivity[36] of the method is low, due to coarse-grained nature of the matching. In case of the latter, the fine-grained matching enjoys a high productivity[37], but all the burden comes to rest on *composition* of the smallest parts - an endless puzzle for linguists.

---

32 IFTT'89 (Oiso) proceedings, p.51.

33 in the mid-1980s; the suggestion was clearly hinting at the high cost of the on-going EUROTRA project

34 the EUROTRA project is a textbook example of this type of approach

35 in large bureaucracies like the EEC administration, it seems to occur that documents or considerable parts of it are retranslated because a previous translation cannot be found or is not known to exist

36 i.e. in real-life MT, it will happen occasionally but not predominantly that whole sentences or larger units can simply be retrieved from previous translations

37 well in accordance with the fact that formal linguists like to deal with the *vastness* of natural language, the *infinity* of its different expressions etc.; a professional translator, on the other hand, and certainly a translator of non-fiction, is much more concerned with convention, text cohesion, consistent use of terminology etc.

In the ABMT paradigm, the size of the unit of matching is inspired[38] by the practice of human translation professionals: sometimes a word, occasionally a sentence[39], but most often something in between: a phrase, a collocation, a word combination, a bunsetsu. The reader is also reminded to Nagao's 1984 seminal paper (see section 2.3 above): the translation is from a block of words to a block of words.

Besides the questions of unit size, a crucial thing is the flexibility of pattern matching: what we want in ABMT is not only to find *exact* matches between units, but also *approximate* ones - just like a professional translator's ability. In BSO's methodology, the painful question[40] "What is approximate matching?" is well solved by the operationally (on the basis of BKB statistics) defined semantic distance, invented by Sadler[41].

ABMT, according to BSO's and certainly to the present author's views, is in between of the two extremes of fig. 3. It is reasonably productive, and relies on a moderate amount of composition. It pays tribute to the principle of *uncompositionality*, rather than to the compositionality of language translations[42]. It means emulating a (junior) translator, not a linguist! This mimicking of translator's skill instead of relying on linguistic formalisms has far-reaching consequences for MT's basic technology. It should be noted that:

- *translator's skill* has never been explicitized in a formal discipline (it is very hard to look in the translators's brain, in terms of expert system knowledge acquisition); however, this situation is an advantage rather than a disadvantage; the present author has no confidence in any efforts[43] towards a formal (rule-based) science of translation;

- instead, recorded translator's output provides a *training set* for ABMT systems;

- the training period (preparation, processing and extension of the training set) will be considerable; one has to think in terms of years, for sure; it will occupy a battery of specially trained translator-type-of-persons equipped with powerful workstations with dedicated BKB-directed software tools;

- a <u>huge</u> translation bank (BKB) is necesssary, because it has to cover all words and terms, and every polysemous word repeatedly[44].

---

38 Canada's translation authority Brian Harris, by his publications on BI-TEXT (a quick translation retrieving tool for professional translators) has done much to convey this inspiration

39 the size of the unit of matching is not *preset, but varies from sentence to sentence* position; and always, the principle of the 'longest match' applies.

40 this question can be rephrased as: What are synonyms? What are near-synonyms? What are related words? etc.

41 a description of this is outside the scope of this report; see the seminar documentation of "MT based on Analogy", Toon Witkam, July 26&27, ATR (part 2-A); or see Sadler, "Working with Analogical Semantics", FORIS 1989, pp. 55-67.

42 see also Nagao's remarks on compositionality in "Machine Translation - How Far Can It GO?" ,'89, p. 64.

43 as for example advocated by Landsbergen (Philips' Rosetta project), University of Utrecht '89; see also interview in Electric Word, March/April '90, p.17.

44 at least 50 times, according to a lexicographic rule-of-thumb

The latter two points seem to confirm that there is no easy or cheap solution to the general problem of MT. However, the reader should be reminded that RBMT, after decades of worldwide efforts, offers no solution at all - whether cheap or expensive. ABMT, in BSO's BKB concretization, comes near to a blueprint for manufacturing MT systems characterized by:

- high-quality output (such as will meet professional translation standards);

- machine learning, effective throughout the ABMT system's operational lifetime.

To conclude this section, it should be mentioned that BSO perceives still a long future for interactive disambiguation. ABMT operation, despite intensive pattern matching backed by a huge and densely structured knowledge bank, will be semi-automatic in most applications. The human contribution will be arranged as a (system-initiated) interaction, which deals with interpretation selections of the source language only. The machine-learning effect exists by virtue of this human factor in the process cycle,

Nagao, 1984:	BSO, 1990:
variety of loose examples	connected text
judgment of human teacher (acceptable / not acceptable)	metric of semantic distances defined over real-life corpus
replaceability of corresponding words is tested by tracing the thesaurus relations	no (explicit) thesaurus used
(mainstream) case frames maintained for verbs	no (explicit) case frames
model of the human brain would be necessary	instead, bilingual corpora will suffice

Table 1. Some differences between Nagao's seminal ideas and BSO's current design.

2.5. Experiments with a BKB simulation model

In the framework of its research on ABMT and the DLT project, BSO[45] has recently built a modest model of a BKB. It comprises 1500 sentences (20.000 words) of text in English, and its (human) translations into French and Esperanto[46]. The text is Chapter 4 of DEC's PC ALL-IN-1 user manual. These three parallel mini-corpora have been processed[47] in accordance with Steps I-IV (as listed in Section 4), in such a way that two mini-BKBs, one for English-Esperanto and one for Esperanto-French were built.

It should be noted that no operational system has been built with these BKBs: there is a (model) engine, but no (model) car as yet. The purpose and scope of the experiment is merely to get more insight into the fabrication process of a BKB, and into its internal structure.

Apart from the software required for months-long (on-screen, interactive, graphically) parsing, synsemizing etc., BSO's software includes a set of access windows to inspect the contents of the resulting model BKBs. With this retrieval interface, the ABMT researcher can quickly[48] get an impression of:

- the *basic representation* used throughout the database, i.e. the word-dependency trees;
- the *valency patterns* below a given word: this amounts to a very rich and flexible set of case frames, here not based on a priori slot designations but on real-life text;
- the different word categories or syntactic functions below which a given input word occurs; all this includes *statistical information*;
- the matches for a given *syntactic pattern* (as opposed to a linear string pattern!); the input pattern may contain wildcards for undefined words;
- *referential relations* (discourse structure): conceptual identity, contrast, inclusion;
- correponding sentences and *translation units* in the other language; back translations and onward translations (English-Esperanto-French transitivity is achieved by coupling the two model BKBs); the principle of full BKB *reversibility*.

The semi-automatic parsing during the build-up phase was driven by simple rule-based parsers, for each of the three languages concerned. Meanwhile, BSO[49] has built a rule-less parser, which uses the already completed part of a BKB for parsing further sentences to enlarge it. The rule-based parsers will therefore only remain useful as booting parsers, to build up an initial volume.

---

[45] in cooperation with DEC

[46] the use of Esperanto is not essential

[47] the volume of this work (including software construction) amounted to 48 person-months; it took place during the period July '89 - May '90

[48] from a DEC3100 workstation

[49] see: Job M. van Zuijlen, "Notes on a Probabilistic Parsing Experiment", BSO/Language Systems, July 1990

3. Application to dialogue MT

In this section, we discuss the distinction 'dialogue vs. text'. We do this with a critical mind, and aware of the 'moving target' of electronic and optical communication appliances over the turn of the century.

We observe the need for a differentiation between dialogue types, and a distinction of information types within a dialogue. We will argue that one type is likely to be handled more appropriately by ABMT than by formalism-based methods because of its strong similarity with written texts.

Finally, the potential of the analogy-based text paradigm for non-MT applications is stipulated. This is relevant because part of what we simply call 'dialogue MT' turns out not to be MT in the strict sense, but rather inferencing and information retrieval with the help of textual forms.

3.1. Dialogues vs. text, a useful distinction?

The paramount question in the context of ATR goals is of course: can ABMT be applied to dialogue translation? After having dealt with ABMT's feasibility for text translation[50], we therefore have to ask ourselves: What distinguishes dialogue translation from text translation?

An answer has already been sketched in a notable COLING'88 paper by Tsujii and Nagao [51]. The main differences they indicated were the clear definition of information and the active speaker and hearer participation in dialogue translation.

It should be noted, however, that the paper was inspired by and commented on ATR's well-known test domains of conference registration and travel agency calls. The authors considered this as an example for a whole class of *goal-oriented* throw-away dialogues, with the following properties:

- the domain is sufficiently restricted and the goal is suffiently clear to justify an (application-oriented) *internal meaning representation* from which a paraphrase (in the other language) rather then a MT-type translation can then be made;

- only for those parts of the input string that appear to be irrelevant to the domain (or goal), conventional syntacto-semantic MT techniques may be applied.

Although we agree with much of the views in this paper, we think it is important to envisage also the patterns of usage of communication and information systems in the

---

[50] in section 2, and also in the 2-day seminar (by the present author) "MT based on Analogy", July 26-27, ATR,Kyoto

[51] Tsujii,Nagao, "Dialogue Translation vs. Text Translation - Interpretation Based Approach", COLING'88, pp. 688-693

next century[52], including future applications of interpreting telephony. Our feeling is that the routine-type-of-dialogues that are very goal-oriented and in a narrow domain, such as:

- inquiring/registering for a conference
- inquiring/booking with a travel agency
- seat reservations (theatre, concert hall, ...)
- car repair/maintenance bookings
- bank transactions

will, in due time (already in the year 2005) be handled by a *man-machine interface*, instead of by man-man communication. These 'semi-mechanic' dialogues will exploit other information channels (visual, pointing)[53], in addition to language.
We will come back to this issue in section 3.1.2.

Another factor is that the increased use of on-screen information tends to gradually blur the distinction between texts and dialogues. This trend will even be reinforced, by the appearance of multimedia and hypertext systems.
One of the biggest markets for multilingual text MT is said[54] to be the translation of (hi-tech) product or equipment documentation. One need only think of the piles of technical manuals that go with a mainframe computer, a process plant or an aircraft. Traditionally, this type of thing has been considered as textual MT. This conception might change however, as not only the storage and presentation, but also the retrieval (and updating) of all these technical data will increasingly be organized electronically. A technician doing aircraft maintenance will look at a monitor screen and obtain the information he wants by pointing to menu items or by asking (speech input), instead of the classical acts of turning pages, consulting index and contents list, following references to other volumes etc. Also he will get more and more of his information in the form of high-definition graphics (with zoom function, of course), though it is unlikely that the use of language will completely disappear.

Under these circumstances, we can expect a proliferation of man-machine dialogues in areas traditionally associated with textual documentation. Only a small subset of these is likely to require international telephone links, but a considerable subset will -within the next twenty years- be equipped with multilingual capacity. This means that *MT technology as a whole will shift across the text-dialogue boundary zone*. It also means that a long-term research programme for basic MT technology cannot afford to address one side (whether text or dialogue) in isolation.

Fortunately, ATR with its unique combination of laboratories is excellently positioned to anticipate future developments in man-machine interfaces or audio-visual techniques, and its consequences for MT.

---

52 on the premise that ATR's developing technology is intended to drive applications at least one decade in the future

53 Cf. work in other ATR laboratories, at MIT's Media Lab (Negroponte) and in Saarbrücken (Prof. W. Wahlster).

54 "Scoping the Prospects for DLT", BSO/McKinsey&Company, 16 December 1988 (confidential)

3.1.1. Partitioning of dialogues (goal-directed vs. irrelevant)

The key concept in Tsujii and Nagao's COLING'88 paper is the *internal meaning representation*. The authors state that in (non-MT) NLU[55] systems, *meaning* can only be defined relative to the internal tasks and processings a system is expected to perform: deductive inferencing, problem solving in some restricted domain, data base access etc.

They then continue to argue that MT, *by itself* (i.e. framed as a mere source-to-target-text conversion, without any further specification of the intended use or purpose of the text) does not define what the internal meaning of a text is and how it can be represented. This deficiency, they claim, lies at the basis of the difficulties peculiar to MT and make it inherently *structure bound.*
They are right, of course, as far as RBMT is concerned[56]. But as soon as we switch to the ABMT paradigm (and its elaboration by BSO), the situation changes: as described by Sadler[57] and supported by Schubert[58], the gist of BSO's ABMT methodology is: *leaving meaning implicit.* In other words: ABMT mimicks a translator, and a human translator does not bother about deep case structures or other formalisms of meaning extraction - he translates on the basis of experience and analogy (both of which come down to *memory*). ABMT, like RBMT, does not have an internal meaning representation. But in contrast to RBMT, ABMT does not need one.

Tsujii and Nagao (COLING'88) propose a layered approach to dialogue MT: internal meaning representation, for the important or relevant parts of the goal-directed dialogue, and conventional MT for its irrelevant or "less important" parts. This implies splitting the dialogue, which we have expressed in fig. 4.

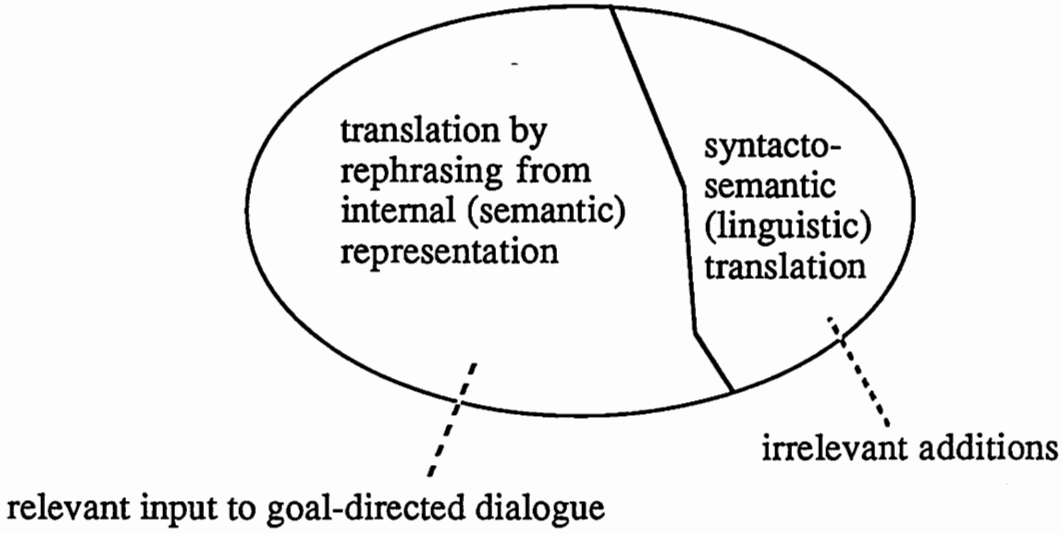

Fig. 4. Partitioning of a dialogue according to processing paradigm.

---

55 NLU = natural-language understanding

56 another author who sharply illuminated this matter is Yoshihiko Nitta: "Idiosyncratic Gap: A Tough Problem to Structure-bound Machine Translation", COLING'86, pp. 107-111

57 Sadler, "Working with Analogical Semantics", FORIS 1989, in particular pp. 11,50,51

58 Schubert, "Implicitness as a guiding principle in machine translation", COLING'88, pp. 599-601

Given this differentiation within dialogues, the question of ABMT's feasibility can now be investigated for each of the partitions separately.

At first sight, the linguistic-translation or "irrelevant" partition seems the closest candidate for ABMT introduction; we will deal with it in section 3.2. (It is tempting for systems designers, to pay little attention to dialogue fragments labelled as 'irrelevant': they could be handled as noise in the input string, not to be further processed, a principle often applied in natural-language DB interfaces. Section 3.1.2 will give counter-arguments.)

But also for the goal-directed partition, ABMT might have something to offer, even when this concerns non-MT (section 3.3).
A potential danger, when working with two completely different paradigms, is of course the possible fuzziness of the partition boundary (fig. 4). A quote from Tsujii,Nagao[59]:

> "The translation system knows in advance what kinds of *information* or *concepts* are important... and also knows a set of surface linguistic expressions which may convey such *important information.*"

suggests that the 'goal-directed' part of the system determines the boundary. We will come back on this in 3.3.

3.1.2. Purposeless dialogues, big future!

After having looked at differences *within* a dialogue, we now take up the differentiation *between* dialogue types. As Tsujii,Nagao[60] point out, there are dialogue types which cannot be handled in the goal-directed mode at all. Their example, "chatterings among housewives without any purposes", is very clear but seems to suggest that the category of "arbitrary" dialogues is not worth serious attention in high technology.

We might, however, make a big mistake by aiming exclusively on the "goal-directed" dialogue type. Developing basic technology for telephone interpretation, as ATR's mission is, there are good arguments for anticipating a growing importance of "arbitrary" or "purposeless" international telephone dialogues:

- the added value of person-to-person communication (compared to person-to-computer) lies in the *fuzziness and subjectivity* allowed by the human use of natural language; beyond 2005, business people will *call* each other when *expressing an opinion* is important; for unpersonal exchange of data, they will turn to their personal computers, and let them handle the exchange, in night time tariff;

- telephony (whether or not combined with a picturephone) will remain an excellent medium for person-to-person communication, i.e. for unrestricted human language.

---

[59] COLING'88
[60] ibid.

The sometimes seemingly purposeless nature of a dialogue, its fuzziness with regard to (artificial) domain boundaries - it is all part of unrestricted, unedited human language, for which the telephone is very appropriate.

In fig. 5 we have sketched what is likely to happen. In the upper half of the figure, we have the goal-oriented dialogues, for which Tsujii,Nagao[61] stated the feasibility of an internal meaning representation (bound to the narrow domain in which the dialogue is known to take place). The emphasize here is on paraphrasing rather than (syntax-preserving) translation. The goal-oriented dialogues will be man-machine after (say) 15 years from now. Tsujii himself is contributing to this shift by his "translation-without-source-text" paradigm[62], in which basically the system (equipped with language and domain knowledge) gathers the information from the human user, by asking questions. Obviously, the intelligence of the system must make up for the lost speaker-hearer cooperation then.

At present, and for the next few years, the goal-oriented dialogues are still man-to-man (customer-clerk or otherwise). ATR's present research effort along the line of illocutionary analysis, intention extraction, plan inference etc. is directed to tackle these dialogues. Speaker-hearer cooperation is an element inherent to this line of research.

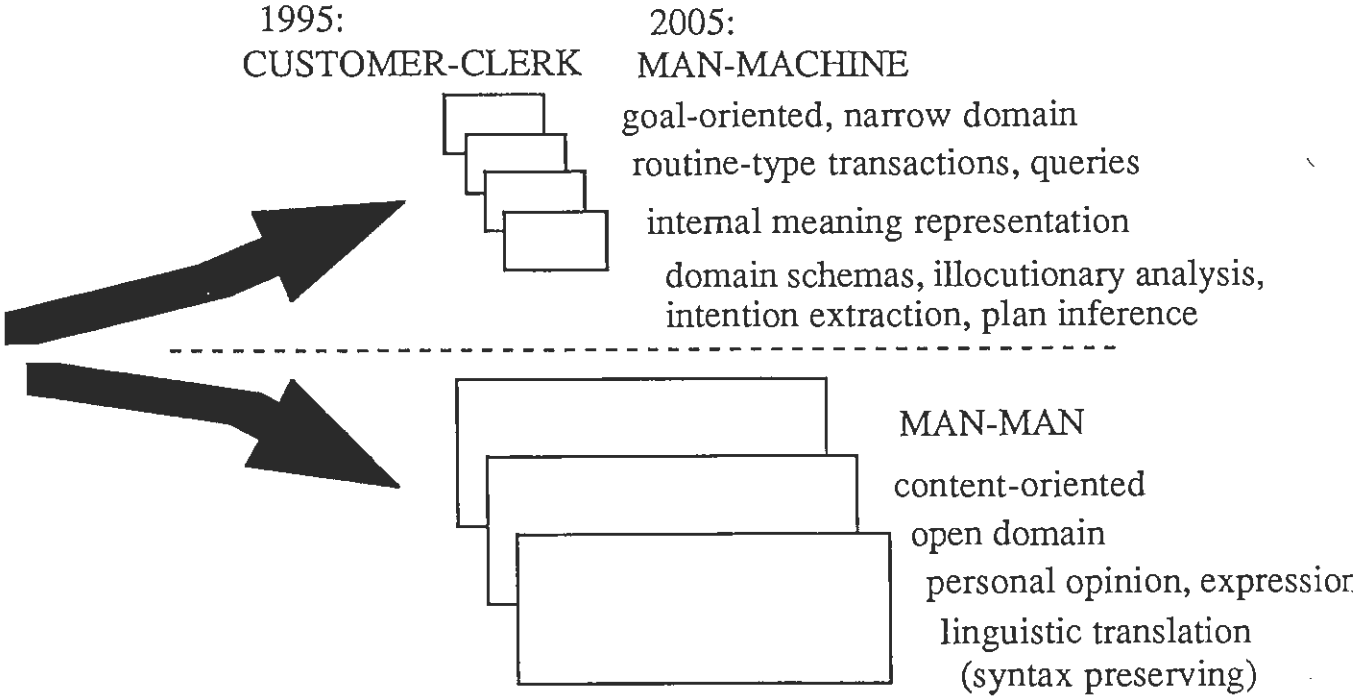

Fig. 5. Differentiation and evolution of dialogues applications.

61 ibid.

62 IFTT'89 (Oiso), p. 50; repeated by Somers in ASTI'89, p. 6-2-2

The other type of dialogues, the so-called 'arbitrary' or 'purposeless', but more appropriately termed 'content-oriented' ones, will remain man-to-man for many decades (maybe ELIZA-style gadgets will be forthcoming that entertain you with a purposeless conversation, but such a thing is unlikely to take place over international telephone lines). These dialogues will be handled by *translation* in the traditional sense, not by paraphrasing. Therefore, syntax as well as semantics must be preserved, according to the 'textbook' paradigm of human translation and the ABMT method which mimicks it. Here, where there is no domain-based inference system to guide the conversation, cooperative speaker-hearer participation will highly contribute to the cohesion and progress of the (semi-automatically translated) dialogue.

The arguments given in favor of "purposeless" dialogues also apply to the "irrelevant" parts of goal-oriented man-man dialogues. The example dialogue in the Appendix, between a customer and a JTB[63] clerk, demonstrates that a goal-directed[64], essentially unpersonal dialogue may include information packed in subtle, subjective wording[65]. The expression of subjective feelings and some introductory chattering is of course quite natural and often essential even in dialogues that are very goal-directed. Just some examples of international consultation:

- enquiries on interest for joint projects (MITI, ESPRIT,...)
- enquiries on available job skills
- exchange of experience between software users
- rapid exchange of information between scientists (e.g. 'cold fusion'!)
- technical aid in urgent situations (oil fire, gas leak, pollution accidents)

Moreover, the goal-directedness of some of these dialogue types may be difficult to abstract and to formalize into an a priori model.

3.2. Dialogues vs. text, technical differences

Going into a bit of detail, the differences can briefly be summarized as follows:

- dialogues have shorter, less wordy or complicated sentences than documents;
- dialogues have more cases of ellipsis and 'ungrammatical' constructions.

---

63 JTB = Japan Travel Bureau

64 the goal, in this particular case, is to file a complaint, to give a bad note to something

65 which, it appears, will be *orally* passed by the clerk to his boss and to the hotel the customer complained about

The latter is related with the former, and also with the presence and 'active participation' (as termed by Tsujii,Nagao[66]) of both speaker and hearer. Note, however, that the above summarized differences are a matter of *degree*. This also applies to 'ungrammatical' constructions: as pointed out by Van Zuijlen[67] and others[68], there is a *scale* of grammaticality. Some utterances are simply less grammatical than others, a viewpoint well in accordance with corpus-based (and therefore ABMT) paradigms.

Sadler[69] has shown how the handling of ellipsis is integrated in the BKB approach, BSO's ABMT method. Throughout this method, attention is given to *text coherence*, including the treatment of various forms of anaphoric reference and ellipsis. To repeat Sadler[70]:

> *"Where the BKB concept breaks new ground is in its combination of two separate dimensions: the horizontal dimension of cross-linguistic equivalence and the vertical dimension of text coherence. This two-dimensional structure allows the BKB to represent... ...discourse structure as well."*

We now claim that a methodology suitable for discourse-based MT[71] can handle dialogues as well as text, since a dialogue is just a specific form of discourse. The differences, as said, are quantitative rather than qualitative: more cases of ellipsis etc. But generally speaking, the difference between text and dialogue is not greater than the difference that can exist between various types of text.

The alternation of speakers, of course, need to be taken into account[72] in processing a dialogue and in preparing example dialogue sets in a BKB. It causes referential relations that cross the speaker/hearer boundary, with the trivial cross-connections of personal pronouns in the 1st and 2nd person:

Language 1:	Language 2:
speaker: Can you send me a registration form?	tourokuyoushi wo kochira ni o-okuri itadake masu ka
hearer: I will send it to you.	o-okuri itashi masu

---

66 Tsujii,Nagao, "Dialogue Translation vs. *Text Translation - Interpretation Based Approach*", COLING'88, pp. 688-693

67 J.M. van Zuijlen, "Notes on a Probabilistic Parsing Experiment", BSO/Language Systems, July 1990

68 e.g. Tomita, ASTI'89

69 see Sadler, "Working with Analogical Semantics", FORIS 1989, pp. 141,142,215

70 ibid., p. 135

71 in contrast to sentence-based MT, which is still predominant; note that BSO gave special attention to text-based MT (see also: Papegaaij,Schubert: "Text Coherence in Translation", FORIS 1988).

72 Sumita has suggested an intermediate level (between sentence level and discourse level) of pairs of word-dependency trees for speaker and hearer

The problem of pronominal ellipsis (including 'zero pronouns') can be addressed by structuring and inserting such examples into a BKB. If we first look for a moment only at Language 1 in the above example[73], the knowledge base will infer the you-I and me-you cross-correlations around the governing verb send by the same probabilistic inference mechanism that Sadler describes[74] for rules like X causes Y => Y due to X, X contains Y => Y in X etc. A similar case would be X sells to Y => Y buys from X.
All this is part of the 'vertical' dimension of the BKB.

Now, looking at the bilingual whole, the 'horizontal' BKB dimension links two languages TU-wise[75]. In other words: with a (highly elliptic) Japanese utterance, there is already connected (via BKB matching[76]) an English 'shadow' utterance, which totally circumvents the formalistic problem of 'zero pronoun resolution'.

This approach can loosely be compared to Kudo's 'local cohesive knowledge'[77], which also addresses the problem of rightly interpreting anaphoric and elliptic forms. The similarity is in the preparatory data collection and the largely automatic building of 'skeletons' (Kudo) or 'microcontexts' (BSO) respectively. Kudo's work appears to us as an ingenuous case frame extension in which the frame covers a question-answer pair. BSO's method is in fact a very rich and flexible case frame method, relying on the semantics-oriented word-dependency representation structure throughout the BKB. The same advantages[78] (compared to case frames) that are claimed for its use in non-dialogue applications are valid for the extension to dialogues.

Adherence to Wahlster's dual user model

Wahlster[79] stated that *bidirectionality* of linguistic knowledge sources is an important design goal for high-quality dialogue translation systems. In combination with this, he stated the need for a *dual* user model. Although this model is much in terms of belief, goals, intended effects, anticipated effects etc., we observe a parallel in the *duality* of BKB-based translation: at the one side, the ABMT system occupies itself with contextual interpretation of the SL, at the other side the system is dedicated to maintain consistency and (stylistic) coherence of the developing TL text[80].
As in Wahlster's scheme, SL analysis and TL generation take place incrementally, with revision and withdrawal of earlier assumptions when appropriate. All the time, there is the cross-connection via (a large diversity of) TUs. The *reversibility* of the BKB[81] is a firm basis now for a continuing alternation of the SL and TL roles; the

---

73 a more elaborate example is contained in the Appendix

74 Sadler, pp. 227-233

75 TU=Translation Unit; delineation of TUs is a semi-automatic process that takes place during BKB construction

76 notice that the flexibility of recognizing semantic proximity is a key principle of BKB matching (i.e. alternation of the verb *send* by the verb *mail* would cause no problem); the BKB design (e.g. Sadler pp. 209...) largely addresses the problems discussed in: Nogaito & Iida, "Noun Phrase Identification in Dialogue and its Application, 2nd Int'l Conf. on ...MT, CMU'88.

77 Kudo, "Local Cohesive Knowledge for A Dialogue Machine Translation System", COLING'90

78 for a summary of basic advantages, see Sadler p. 144-145

79 Wahlster, "Anticipation Feedback in Dialog Systems", ASTI'89

80 see Sadler, pp. 159-195

81 ibid. p. 144

scheme below indicates the principle (DB = database, here used for one language half of the BKB):

textual MT:
source and target language do not change:

SL = Language A; interpret SL expression correctly (i.e. find close examples in SL DB; reliable SL reading)	TL = Language B; form TL expression correctly (i.e. close to examples in TL DB; TL consistency, good TL style)

dialogue MT:
source and target language alternate:

SL = Language A; interpret A expression correctly (i.e. find close examples in A DB; reliable A reading)	TL = Language B; form B expression correctly (i.e. close to examples in B DB; B consistency, good B style)
TL = Language A; form A expression correctly (i.e. close to examples in A DB; A consistency, good A style: not only with respect to DB, but also with respect to current input text	SL = Language B; interpret B expression correctly (i.e. find close examples in B DB; reliable B reading)

The treatment of dialogues requires some special attention to the way discourse consistency is implemented. For instance, across the speaker-hearer 'boundary', consistency should be such that in case of (near-)synonyms, the one introduced first will consistently be used by the other speaker (at least in a number of 'dialogue models'). So suppose *brochure* and *pamphlet* are synonyms, then if speaker A first used *pamphlet*, the system should maintain this word use in its translation of speaker B, even if the default choice (on statistical preference or whatever) were *brochure*.

A particular constraint of dialogue translation, certainly in comparison to batch translation of written text, is its forcing real-time character. There is no way to postpone translation beyond the next speaker/hearer role reversion, and these reversion points may be just one sentence apart. This seems in pitiful contrast to the current trend[82] from sentence-based to text-based translation. However, the limitation is only in the "downstream" (look-ahead), not in the "upstream" (look-back) direction: the cohesiveness of the dialogue text can therefore at least half be utilized for finding cues to resolve ambiguities. Moreover, remaining problems can be dealt with in a *metadialogue* with the speaker. This is, in fact, not different from the strategy envisaged for BSO's DLT system[83], in which a disambiguation dialogue[84] intervenes after processing of (in principle) each sentence.

---

82 viz. gradually increasing popularity of corpus-based MT, including BKB and Laffling (footnote 26)

83 see Sadler p. 150 (DLT = Distributed Language Translation)

84 because the intended application of DLT was not *dialogues*, there was no direct need to call the disambiguation dialogue a *metadialogue*

3.3. Non-MT application of the analogy-based paradigm

AS argued by its inventor[85] (and fully backed by the present author), the BKB is a sound basis for making inferences. These need not necessarily be used for translation: they can also support various other applications, such as text management (detection of inconsistencies in technical documentation!), concept-oriented (instead of string-oriented) information retrieval, abstracting etc.

Coming back to the dialogue partitioning of 3.1.1, the fuzziness of the boundary makes it dangerous to have incompatible methods operating at each side. The fact that BKB-style ABMT works fully concept-oriented and favors probabilistic and dynamic[86] formation of inference rules, makes it an attractive candidate for the goal-oriented dialogue part as well. Also the determination of the boundary itself, for which Tsujii,Nagao[87] appear to envisage surface strings (our underlining):

> "The translation system knows in advance what kinds of *information* or *concepts* are important... and also knows a set of surface linguistic expressions which may convey such *important information.*"

can be done more robustly by a concept-based instead of a string-based method. It goes without saying that superiority of the former will only be realized if there it can rely on a large collection of example data: dialogues (representative for those to be handled by the ABMT system), manually translated, checked and preprocessed into BKB form.

Considering the fact that ATR is in the process of preparing basic technology with which implementors can address a variety of different goal-directed domains in the near future, the 'boundary problem' can become overwhelming if not met by a robust, integral solution. One needs to anticipate a system that can handle the co-existence of many different domains, with very fuzzy boundaries. The American CYC project[88] is aimed at such a goal, and so will large-scale BKB construction.

---

85 Sadler pp. 209-233, 241-243

86 'dynamic' refers to the mechanism of BKB extension by ABMT production output; it is also connected to the notions of 'recency' (the system can prioritize information from recent BKB portions) and 'learning' (from new input); see also Sadler, p. 131

87 COLING'88

88 at MCC, Texas (see CACM, August 1990)

4. Basic technology (MT manufacturing) considerations

ATR's mission is to develop *basic technology* for interpreting telephony; with this as a basis, Japanese (or foreign[89]) industry will then develop and commercialize future *applications*.

In order for this rationale to become true, the basic technology needs to include more than the elaboration of an *operating method* for telephone interpretation: it needs to include a *manufacturing method* as well. Maybe for the scientist, the former is the most interesting, but entrepreneurial thinking will tend to focus on the latter: the manufacturing method determines costs, production speed, ease of customization, scope for product diversification etc.

The *basic technology* should therefore by all means include a manufacturing technology, and this should be such that its methodology can be relatively easily[90] moved ('ported') from one application domain to the next. What's needed, in fact, is a *production method*, which enables manufacturers to set up a flexible production line of MT modules for various applications: general ('off-the-shelf') as well as customer-specific ones.

ABMT lends itself well to this requirement because:

- ABMT uses real-life text as its knowledge source; this means that records of a customer's recent text (or dialogues) can be used as examples; *"we turn your past translation work into a BKB, which can then drive a largely automated translation shop for you"* might become the slogan for the service-oriented ABMT vendor of the next decade (see fig. 6);

- the process of ABMT prefabrication, though substantial (see steps I-IV below), does not contain risky elements, i.e. it leaves no room for subjective or arbitrary interpretation decisions; also, it does not depend on theoretical modelling;

- ABMT's raw material is *human translation*: high-quality, professional translation; if necessary, this raw material can be made to order as the first step in manufacturing a BKB (see Step I below); organized human translation (team)work is a familiar activity in business and industry[91];

---

89 part of ATR's research results circulate in the public domain; apart from that, cooperative application development with (or licensing to) foreign companies could occur

90 i.e. without having to redo research or exploratory development

91 see JEIDA Report '89: "A Japanese View of MT in light of the Considerations and Recommendations Reported by ALPAC, USA"; this report includes a survey on human translation practices in Japan; one of its findings which is of interest to ABMT is (p. 8): *"The translation is done sentence by sentence.... Translation by paragraph...is seldom done."* This provides good conditions for alignment of the parallel corpora with regard to BKB construction.

- the process of ABMT prefabrication is semi-automatic: knowledge acquired in the already manufactured part of the BKB is used to preprocess further source material; in this way, the BKB prefabrication is an increasingly automatic process; moreover, the machine learning cycle can go on at the user site, by adding the system's daily production[92] to the BKB contents.

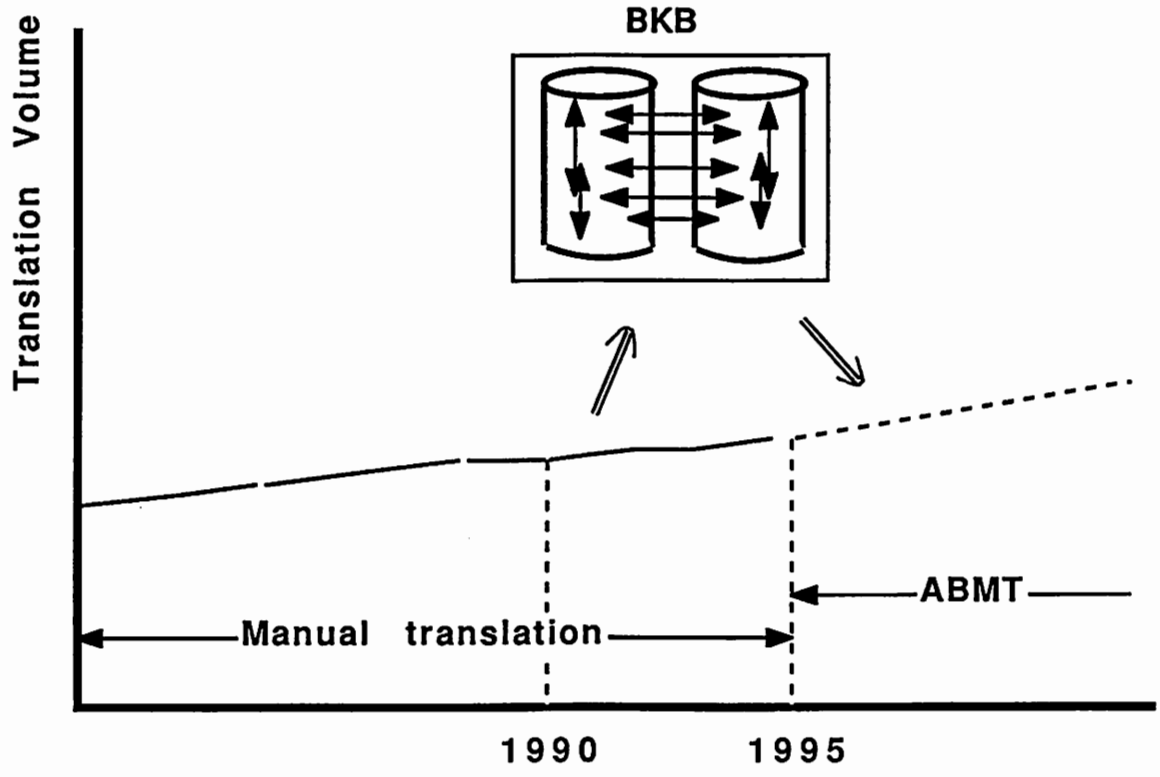

Fig. 6. Recycling of translation work. From the viewpoint of the MT user, it is his own corporate writing style and specific terminology (in both languages) that are fed into the ABMT system.

[92] given certain authorization and security safeguards, to prevent 'pollution' or 'infection' of the customer's BKB

We will now briefly list the elementary steps and issues in BKB manufacturing, which consists of 4 main steps[93]:

I. Corpus selection and rectification
- domain delineation
- balancing
- quality assessment (peer review)
- supplement missing parts
- redo 'skewed' translation
- correct terminological inconsistencies etc.

Step I can be considered as "the translation-bureau" step in the BKB manufacturing process. Qualified professional translators will be involved, equipped with state-of-the-art CAT[94] tools, on-line term banks etc.

II. Syntactic parsing
- string-to-tree conversion
- word-dependency trees
- according to dependency syntax
  (which exists for each language separately)
- semi-automatic
- graphic or linear (color, arrows) visualization
- structural disambiguation
- corresponding sentence string in parallel
  language can be consulted
- parser more powerful as parsed volume grows

Step II should be perceived as a battery of applied linguists at workstations. Though special training is needed, this job will become routine. A ruleless, corpus-based parser is appropriate, though a rule-based ('booting') parser
can be used to acquire an initial volume of parsed-trees (for the ruleless parser).

III. 'Synsemizing'
- identifying TUs
- thereby 'synchronizing' the two corpus halves
  (making the 'horizontal links')
- graphic or bilinear (color, arrows) visualization
- semi-automatic
  (by increasing system recognition of recurring TUs)
- system's suggestions can be overruled
- human can identify interesting translation elements
  ('translation idioms')
- but all words will be covered by interlinked TUs

Step III can very well be done by professional translators. Though special training is needed, this job will become routine. As in Step II, the automatic component will become more powerful as the processed corpus volume grows, thereby gradually reducing the workload on the (human) 'synsemizers'.

---

93 see also: Sadler (pp. 137-142), and: Session 5 of the 2-day seminar (by the present author) "MT based on Analogy", July 26-27, ATR,Kyoto;

94 CAT = Computer Aided Translation; i.e. j-generation MT may be used in the process of building j+1-generation MT

IV. Adding discourse structure
- identifying anaphoric referents
- monolingual, for each corpus half separately
- form of disambiguation[95]
- system can do clever suggestions
- not only anaphora, also identities, inclusions and exclusions

Step IV requires an excellent (human) understanding of the contents of the text as a whole, including the right interpretation of zero pronouns and the interlacing of reference links across the speaker-hearer boundary (for dialogues).

If the intended ABMT application type is *dialogues*, the BKB should of course be equally composed of a large quantity of dialogues. For a narrow domain in a closed user group or controlled environment[96], a corpus size in the order of magnitude of the present ATR database[97] (several hunderd thousand words) may be adequate. If the application is aimed at the general public or an open user group, some relaxation of "speaking standards" and domain boundary is unavoidable, and a tenfold increase in BKB size will soon be necessary.

5. Proposals for the next step

> *"The solution cannot be simple. We need and should process a large amount of real data"* [98]

Because of its drastic deviation from 'mainstream' (rule-based) MT, and because of its reliance on large quantities of data (corpora), ABMT prototyping is not an easy affair. A projection by the present author is 10 MECU[99] for a minimal *production prototype* with a corpus size of 3 million words per language.

Between this projection and the current research status, a number of intermediate steps can be conceived. The 2 proposals below are small steps, departing from the present state (which includes the model described in 2.5). Each of them has a duration of 1 year, and a limited volume of effort[100].

---

95 according to Geoffrey Sampson (Leeds, UK) there is an unsurpassable(85% or so) boundary to automatic resolution of anaphora

96 meant is a situation in which corporate or official instructions and guidelines impose a standard vocabulary and phraseology; an example might be the "Policespeak" reportedly (Electric Word, May/June 1990) being prepared for the English and French police and rescue workers at both sides of the Channel Tunnel

97 Ehara, Ogura & Morimoto: "ATR Dialogue Database", Int'l Conf. on Spoken Language Processing, 18-22 November 1990, Kobe.

98 conclusion presented by K.Shirai in Panel Discussion at ASTI'89

99 MECU = millions of ECU (European Currency Unit)

100 see also: Witkam, "Technical Proposal involving research on ABMT", 31 August 1990

In opting for one (or possibly both) of the proposed continuation steps, ATR will have the following benefits:

- time gain in pursuing an alternative paradigm of great potential (ABMT), by making use of available outside expertise; as argued repeatedly in this report, ABMT is part of basic technology, and ATR cannot afford to miss this development;

- even if a large-scale commitment to ABMT as part of ATR's basic technology would appear to be undesirable, ABMT could still be quite useful[101] as part of a hybrid approach.

As a special consideration, ATR's affinity with speech processing and the importance of big corpora in speech technology make ABMT interesting to pursue: it may hold still hidden benefits for the speech-language interface.

5.1. Adding Japanese to an experimental BKB model

This proposal consists in adding *Japanese* to the already existing English-French BKB simulation model built by BSO[102]. The result will be a Japanese-English BKB whose internal structure can be visualized by a powerful and flexible retrieval mechanism (see 2.5 for features). As a consequence of BKB reversibility, it can be used in two directions (English-Japanese as well).

Apart from studying the BKB's internal (graphically presented) word-dependency tree *representation* and the delineation of TUs[103] for a fair amount and diversity of real-life sentences, the construction of a Japanese BKB-half will involve extensive hands-on exercising in BKB *manufacturing*, in line with Steps I-IV mentioned in Section 4:

I. deskewing
II. semi-automatic parsing
III. synsemizing (TU identification)
IV. discourse structure adding

Each of the above 4 activities is estimated to take a full 2 months for a shrewd (and specially instructed) translator. In addition, 2 months at an early stage are required for finalizing a Japanese grammar rule set (for the rule-based parser); a word-dependency grammar for Japanese has already been written[104], but requires critical revision in view of the here proposed task and the nature of the (DEC) texts. This

---

101 it could for instance be applied to check synonymy in conjunction with other methods, such as Melchuk's lexical functions (see: Stanwood, Suzuki "Some Computational Applications of Lexical Functions", NLC90-12, p. 7)

102 the model is installed on one of ATR's DEC3100 workstations; it includes the tools for BKB manufacturing (albeit of a preprototypical nature)

103 TU = Translation Unit, a basic notion in BKB-style ABMT

104 Shigeru Sato & Shin'ya Noe, "A Dependency Syntax of Japanese", BSO 1988; see also: Maxwell & Schubert, "Metataxis in Practice", FORIS '89, pp.183-206

revision can take place in parallel with Step I; all other activities need to be done consecutively. A fifth step consists in the actual use, for demonstration and study, of the constructed Japanese-English BKB. A project duration of 1 year must be reckoned with.

5.2. Stochastic simulation of ABMT operation

In order to analyze better the potentials of ABMT, not only the manufacturing and internal structure of a BKB need to be studied (as in 1.1), but also its *operation* in translation run-time.

The addition of a run-time module to the existing BKB model (section 2.5) would be a considerable software project. But even if that were achieved, the running of a 20.000-words BKB would not be very representative for studying matching chains and their time and memory usage patterns of a prospective 3-million-words production prototype.

In order to sharpen our view on ABMT operation without having to take a big investment step, we propose *stochastic simulation* of ABMT *operation*. In this, the costly-to-build BKB will be replaced by a *BKB-emulator*, driven by carefully worked out random generators over various predefined functions (incorporating lexicographic and language-statitistical knowledge).

It will have 2 functions:

1. Improving our insight of how full-blown[105] ABMT operates.
ABMT, though its principles are relatively simple, is a quite complicated mechanism when at work at an actual sentence. The flexible pattern recognition for approximate matching of previous TUs, the composition of the sentence out of such (partially overlapping) examples, the resolution of conflicting evidence - all these factors are important but difficult to grasp (as to their *quantitative* as well as qualitative effects) in a paper description of one or two examples[106].
A stochastic simulation with well-visualized (color graphics) output, can contribute to the understanding of overall ABMT functioning and the interdependence of the various factors involved. It will enable researchers to study the effects of changing global BKB constants or averages such as:

- vocabulary size
- corpus size
- repetitiveness of the text type
- idiomaticity
- TU size
- lexical divergence (SL-TL)[107]

---

[105] i.e. ABMT as the *only* available mechanism (in contrast to *hybrid* approaches)

[106] Sadler (see for instance his pp. 159-181) has made a brave effort in paper descriptions

[107] SL = source language, TL = target language

2. Predicting ABMT's time and memory usage patterns. ABMT appears to be computation and memory intensive. A semantic proximity computation can easily cause a cascade of pattern matching operations, not restricted to a local region of the BKB. In conjunction with Function 1 (above), resource usage can be simulated, visually presented and recorded. Moreover, the effect of variables such as *cache size* and a few alternative designs (e.g. *prefabricated lists of semantic proximities* at the higher end of the Zipf curve) can be tested. Also, such a simulation might help to find out to what extent ABMT will gain from forms of parallel processing.

It should be realized that a stochastic simulation as indicated will not give an answer to *all* questions. It will be *supplementary* to other research instruments: detailed on-paper analysis of example sentences, inspecting a small (but real) mini-BKB, etc. Function 1 can be considered as an electronic extension of the ABMT lecturer. For both functions, but most clearly for Function 2, it is indispensable that the package includes full documentation of all data and assumptions (linguistic or otherwise) on which the simulation is based.

Sadler[108] gives a good flavor of what happens in ABMT (and what we would like to simulate):

> *"The whole process should be thought as a spreading ripple of implication passing through the text representation whenever new information comes in..."*

In his honor, the codename 'RIPPLE' is suggested for the proposed simulation project.

---

[108] p. 230, middle

Acknowledgement

The author would like to thank Akira Kurematsu, president of the ATR Interpreting Telephony Research Laboratories, for his encouragement; Eiichiro Sumita for his steady cooperation; and Hitoshi Iida and Terumasa Ehara for their interest shown in this research.

Appendix

On the following 9 pages, a dialogue (from ATR's Dialogue Database) and -from this dialogue- a number of monolingual reference groups, recurrent (bilingual) translation units, and BKB-style dependency fragments (extracted world-knowledge) are displayed.

担当者：ありがとうございます、ＪＴＢ本社内支店でございます。
申込者：杉田といいますけど、申し込みじゃないんですけど。
担当者：はい、どのようなことでしょうか。
申込者：ええ、実は先日おたくのお世話になって、下賀茂温泉に行ってきたんですけど、ちょっといいづらいんですけどね、期待はずれだったんです。
担当者：そうですか。
　こちらのパック旅行にご参加いただいたんですか、それともお宿の手配だけさせていただいたんでしょうか。
申込者：もちろんパックです。
　名前は「下賀茂温泉味自慢舌つづみ」だったと思います。
担当者：はい、いつのご参加だったんですか。
申込者：１０月１９日から１泊２日です。
　友達２人といきました。
担当者：承知いたしました。
　どういった点がご不満だったんでしょうか。
申込者：ええ、一番期待はずれだったのは、お料理でしたね。
　伊勢エビ、平目、さざえとかいろいろ種類はありましたけど、とても、とりたてで新鮮そのものとは、いえないものでした。
　パンフレットの写真とは全然違いましたよ。
担当者：そうですか、パンフレットのお写真は４人前ということですので、多少３名様ですと違った感じにはなるかもしれません。
　食事のせいでご病気になられたとかそういうことではないんですね。
申込者：ええ、別にお料理で食中毒を起こしたわけじゃありません。
　ただ、このパックツアーは、新鮮な海の幸が売り物でしょ。
　そのメインがこれじゃ、と納得できなくてね。
担当者：そうですねえ、今までたくさんの方にこの企画はご利用いただいてるんですが、そうしたことはあまり聞きませんでしたし、また具体的に何か不手際があったということでなければ宿の方にもフィードバックはできないんですけれども。
申込者：ええ、それはわかってますけど。
　でも、ついでに、言わせていただくと、露天風呂なんかも写真とは違って、とても小さく、おまけに眺めも決していいものとはいえませんでした。
担当者：そうですか、お宿の石廊館は、あまり大きな旅館ではありませんし、それほど新しくはありませんが、手入れは行き届いたいいお宿のはずなんですが。
申込者：設備の方は問題はなかったのかもしれないですけど、そこの従業員の態度が少し、サービス精神に欠けていたような気がして感じ悪かったんですよ。
担当者：何か失礼があったんでしょうか。
申込者：特別、嫌なことがあったわけじゃないんですけど、なんていうんでしょうね、無愛想なんですよ。
担当者：そうですか、ただやはりそういったこともなかなか旅館の方には伝えにくいんですね。
申込者：それと、申し込みの時には、できれば眺めのいい部屋を希望したんですけど、実際は眺めなんか全然よくなくって、それにその旅館はほかにもいい部屋が、ずいぶん空いていたような気がしましたよ。
担当者：ただこの企画はお料理が中心ですので、必ずしもそうしたご希望どうりという訳にはいかないんですね、そのあたりもお汲みおきいただきたいんですが。
申込者：そうですか。
担当者：ええ、ですからお料理がまるっきり違っていたとか、従業員が何か粗相をしたとか、そういったことでもあれば、こちらから逆に旅館の方にきつくいうこともいたしますが、具体的になにもないとちょっとそうしたこともできにくいんですね。
　そのために旅館の方に落度がなかったかお伺いした訳でして。
申込者：今さら、あれこれいっても仕方ないですしね。
担当者：はい、ともかく、お客様の方からこういう意見があったというのを上の者ですとか、旅館の方にもお伝えしておきますので云どうかこれからも当社をごひいき下さいますようお願いいたします。
申込者：そうですね、ほかの旅行者のためにもこういう少数派の意見を尊重していただくことは、とても素晴らしいことだと思います。
　また、機会がありましたら、こちらこそよろしくお願いしますよ。
担当者：どうもありがとうございます。
　私どもはサービス業ですので、当然そうしたお客様のご意見は今後の業務に反映させていきたいと考えておりますので。
申込者：わかりました。
担当者：ではこれからもよろしくお願いいたします。
申込者：じゃ、これで失礼します。
担当者：ごめんくださいませ。

	CUSTOMER:	CLERK:
1		Thanks for calling. This is JTB.
2	This is Sugita. I'm not applying.	
3		What can I do for you?
4	I went to Shimogano with your help the other day. And this is diificult to say but I'm afraid I was disappointed.	
5		Were you? Did you join a pack tour or did we only arrange a hotel for you?
6	Of course, I joined one of your tours. I think its name was "Enjoy delicious food in Shimogano".	
7		When was it?
8	From October 19th, it was a one night-two day tour. I went with my friend, so two people in all.	
9		What was most disappointing?
10	What I was most disappointed in was the food. There were many kinds of food such as lobsters, flatfish, and turbot, but they weren't fresh at all. They were quite different from the foods shown in the pamphlet.	
11		Is that so, the pamphlet showed food for four people so it is possible that the foods they gave you were a little different. You didn't get sick because of the food, did you?
12	No, I didn't get poisoned by the foods, but the fresh marine products were the point of the tour, weren't they? I wasn't satisfied with the grade of them.	
13		Well, many people have attended the tour and we haven't heard of such before. We can't say anything to the hotel unless we are absolutely sure that they made a mistake.
14	Yes, I understand that but in addition to the food, the open-air baths were different from the pictures in the pamphlet and were very small and they didn't have a good view.	
15		I see.

		The hotel, Iroukan , isn't a very big and isn't that new, but it must 15 be a well-kept one.
16	The facilities were not bad, but the staff's manner was bad, I mean they lacked the service spirit.	
17		Did they do something rude to you?
18	No, in particular, but, well, what should I say, I mean they seemed to be impolite.	
19		I see. It is hard for us to tell them something abstract like that.
20	And I asked for a room which had a good view when I appliedbut my room didn't have a god view at all and I felt that there were many other good rooms that were empty.	
21		We can't always satisfy such wishes because the main point of the 21 tou is food. We hope you will understand that.
22	Well, yes.	
23		If the food was quite different or the staff did something rude to 23 you, we could complain to them but without something solid, we 23 can't do anything. That's why I asked you about such cases.
24	It doesn't make any sense to complain now, does it?	
25		Anyway, I received an opinion from you, so I will let my boss and 25 the hotel know about it. I hope you will continue to be our customer in the future, too.
26	Yes, I think it is very important to look into the opinions of minorities. If I have another chance to travel, I will ask you.	
27		Thank you.
27		Our work is in the service area so we, of course should utilize your opinion within our business.
28	I see.	
29		We are always ready to serve you.
30	Good bye.	
31		Good bye.

REFERENCE GROUPS:
pamphlet:

REFERENCE GROUPS:
pamphlet:

A-4

tour:

1----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ------
2----- --------- --- -------- ---- --- ---------- --- --- ---------- --
3---------- --- ---------- -- ---------- --- --- ----- -
4--- -- ---- ------ ----- --- ---------- ---- --------- ---- ----- -----
5-------- ---------- -------- ------- ---------- ---------- --------
6-------- ----- ------- --- -------- -- -- -----one of your tours.
6----- --------- its name was --- ---------- --- --- ---------- --
7--- -- it ?
8----- ---- ------ it was a ---------- ---- --------- ---- ----- tour
9----- --------- --- -------- ---- --- ---------- --- --- ---------- --
10---------- --- ---------- -- ---------- --------- -----
10----- ---- ---------- ---------- ---------- --- ---------- -------
10--- ----- --- -- ----- --- ---------- ---- --------- ---- ----- -----
10 ------ -- ---------- ------ ------ ---------- -- ---------- --- --
11---- -- --- ------ ----- --- ---------- ---- --------- -- ----- ---
11------- ---------- -------- ------- ---------- ---- ------ -------
12-- -- ---- ----- --- ----------,------- ---- --- the tour ---
12----- ---- ------ ----- --- ---------- ---- --------- ---- ---- ----
13-------- ----- ------- --- the tour --- ---------- -------- -------- -
14---- -- ---- ------ ----- --- --------- --- ---- --------- ---- ----- ---
15----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ----
16------- --------- --- -------- ---- --- ---------- --- --- ----------
17- -- ---------- --- ---------- -- ---------- --- --- --
18--- ------ ---- ------ ----- --- ---------- ---- --------- ---- -----
19-------- ---------- -------- ------- ---------- ---------- -------
20--------- ----- ------- --- -------- -- -- ---------- -------- --------
21-- -- ---- ------ ----- --- ---------- ---- the tour ---- ----- -----
22----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ----
23------- --------- --- -------- ---- --- ---------- --- --- ----------
24- -- ---------- --- ---------- -- ---------- --- --- --
25--- ------ ---- ------ ----- --- ---------- ---- --------- ---- -----
26-- ---- ------ ----- --- ---------- ---- --------- ---- ----- -------
27-------- ---------- -------- ------- ---------- ---------- -------
28--------- ----- ------- --- -------- -- -- ---------- -------- --------
29----- -- ---- ------ ----- --- ---------- ---- --------- ---- ----- --
30----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ----
31------- --------- --- -------- ---- --- ---------- --- --- ----------

food:

1----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ------
2----- --------- --- -------- ---- --- ---------- --- --- ---------- --
3-------------------- --- ---------- -- ---------- --- --- ----- -
4--- -- ---- ------ ----- --- ---------- ---- --------- ---- ----- -----
5-------- ---------- -------- ------- ---------- ---------- --------
6-------- ----- ------- --- -------- -- "------ ------- food -- -------"
7--- -- ---- ------ ----- --- ---------- ---- --------- ---- ----- -----
8----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ------
9----- --------- --- -------- ---- --- ---------- --- --- ---------- --
10-------------------- --- ---------- -- ---------- the food -----
10----- ---- many kinds of food such as lobsters, flatfish and turbot
10--- they --- -- ----- --- ---------- ---- --------- ---- ----- -----
10 They- -- ----------quite different from the foods-------- --- --
11---- -- food ------ ----- --- ---------- ---- --------- -- foods ---
11------- ---------- -------- ------- ---------- food ------ -------
12-- -- ---- foods --- the fresh marine products ---- ----- ---- -- ---
12----- ---- ------ ----- --- ----------~ ---- --------- ---- ---- them
13-------- ----- ------- --- -------- -- -- ---------- -------- -------- -
14---- -- ---- ------ ----- --- the food --- ---- --------- ---- ----- ---
15----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ----
16------- --------- --- -------- ---- --- ---------- --- --- ----------
17- -- ---------- --- ---------- -- ---------- --- --- --
18--- ------ ---- ------ ----- --- ---------- ---- --------- ---- -----
19-------- ---------- -------- ------- ---------- ---------- -------
20--------- ----- ------- --- -------- -- -- ---------- -------- --------
21-- -- ---- ------ ----- --- ---------- ---- --------- ---- ----- food
22----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ----
23--- the food---- --- -------- ---- --- ---------- --- --- ----------
24- -- ---------- --- ---------- -- ---------- --- --- --
25--- ------ ---- ------ ----- --- ---------- ---- --------- ---- -----
26-- ---- ------ ----- --- ---------- ---- --------- ---- ----- -------
27-------- ---------- -------- ------- ---------- ---------- -------
28--------- ----- ------- --- -------- -- -- ---------- -------- --------
29----- -- ---- ------ ----- --- ---------- ---- --------- ---- ----- --
30----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ----
31------- --------- --- -------- ---- --- ---------- --- --- ----------

A-6

hotel,facilities, staff, room:

1----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ------
2----- --------- --- -------- ---- --- ---------- --- --- ---------- --
3---------- --- ---------- -- ---------- --- --- ----- -
4--- -- ---- ------ ----- --- ---------- ---- --------- ---- ----- -----
5-------- ---------- -------- ------- ---------- a hotel --- --------
6-------- ----- ------- --- -------- -- -- ----- ---------- ----------
6----- --------- ---------- --- ---------- --- --- ---------- --
7--- -- ----------
8----- ---- ------ -- --- -- ---------- ---- --------- ---- ----- -----
9----- --------- --- -------- ---- --- ---------- --- --- ---------- --
10---------- --- ---------- -- ---------- --------- -----
10----- ---- ---------- ---------- ---------- --- --------- -------
10--- ----- --- -- ----- --- ---------- ---- --------- ---- ----- -----
10 ------ -- ---------- ------ ------ ---------- -- ---------- --- --
11---- -- --- ------ ----- --- ---------- ---- --------- -- ----- ---
11------- ---------- -------- ------- ---------- ---- ------ -------
12-- -- ---- ----- --- ---------- ---- --- -------- ---
12----- ---- ------ ----- --- ---------- ---- --------- ---- ---- ----
13-------- ----- ------- --- -------- --- the hotel -------- -------- -
14---- -- ---- ------ ----- --- --------- --- open-air baths ------- ---
15-- ----
15 The hotel --- ---- --------- --- ---------- ---- --------- ---- ----
16 The facilities ------ ----- --- -------- staff --- ----------they-----
17--- they ---------- --- ---------- -- ---------- --- ---
18--- ------ ---- ------ ----- --- ---------- they --------- ---- ----
19-------- ---------- -------- ------- them ---- ---------- -------
20--------- a room -------- --- -------- -- -- my room ---------- ---
20----- -------- ---------- many other good rooms ---- ---- -----
21-- -- ---- ------ ----- --- ---------- ---- -------- ---- ----- -----
22----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ----
23------- --------- --- -------- the staff---------- --- --- ----------
24- -- ---------- --- ---------- -- ---------- --- --- --
25--- ------ ---- ------ ----- --- ---------- ---- --------- ---- -----
26-- ---- ------ ----- --- ---------- ---- --------- ---- ----- -------
27-------- ---------- -------- ------- ---------- ---------- -------
28--------- ----- ------- --- -------- -- -- ---------- -------- --------
29----- -- ---- ------ ----- --- ---------- ---- --------- ---- ----- --
30----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ----
31------- --------- --- -------- ---- --- ---------- --- --- ----------

hotel, facilities, staff, room:

1
2
3
4
5お宿
6
6
7
8
9
10
10
10
10
11
11
12
12
13宿
14露天風呂
15お宿旅館お宿
16設備そこ の 従業員
17(ø は)
18
19旅館
20部屋旅館いい 部屋
20
21
22
23従業員旅館旅館
24
25
26
27
28
29
30
31

WORLD KNOWLEDGE:

1----- ---- ------- ----- --- ---------- ---- --------- ---- ----- ------
2----- --------- --- -------- ---- --- ---------- --- --- ---------- --
3-------------------- --- ---------- -- ---------- --- --- ----- -
4--- -- go to Shimogamo-- --- ---------- ---- --------- ---- ----- --
5-------- ---------- -------- ------- ---------- ---------- --------
6-------- ----- ------- --- -------- -- -- ----- ---------- ----------
6----- --------- enjoy food----- --- -------delicious food----- --- --- -
7--- -- ----------
8----- ---- ---go with friend---------- ---- --------- ---- ----- ----
9----- --------- --- -------- ---- --- ---------- --- --- ---------- --
10-------------------- --- --------disappointed in food------ -----
10----- ---- ---------- ---------- ---------- --- ---------- -------
10--- ----- --- -- ----- --- ---------- ---- --------- ---- ----- -----
10 ------ -- ---------- ------ ------ -----foods shown in pamphlet
11---- -- --- ------pamphlet shows food ----- --- ---------- ----
11------- sick because of food -------- ------- ------- ------ -------
12-- -- ---- poisoned by foods -------------------- ---- --- --------
12----- ---- ------ ----- --- --------- fresh marine products --------
12------ satisfied with the grade ----- --- grade of marine products
13---------- --- ---------- -- ---------- --- --- -----
14---- -- ---- pictures in pamphlet --- ---- --------- ---- ----- ---
15----- -big hotel--- ------ new hotel----- --- -----well-kept hotel
16---bad facilities --------manner of staff-------bad manner --------
17-- ---------staff lack---------- lack spirit------- service spirit----
18--- ------ ---- ------ ----- --- ---------- ---- --------- ---- -----
19-------- ---------- -------- ------- ---------- ---------- -------
20--------- ----- ------- --- -------- -- -- ---------- -------- --------
21-- -- ---- ------ ----- --- ---------- ---- -------- ---- ----- -----
22----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ----
23------- --------- --- -------- ---- --- ---------- --- --- ----------
24- -- ---------- --- ---------- -- ---------- --- --- --
25--- ------ ---- ------ ----- --- ---------- ---- --------- ---- -----
26-- ---- ------ ----- --- ---------- ---- --------- ---- ----- -------
27-------- ---------- -------- ------- ---------- ---------- -------
28--------- ----- ------- --- -------- -- -- ---------- -------- --------
29----- -- ---- ------ ----- --- ---------- ---- --------- ---- ----- --
30----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ----
31------- --------- --- -------- ---- --- ---------- --- --- ----------

WORLD KNOWLDGE:

1·····
2·····
3·····
4·····下賀茂温泉に 行く·····
5·····
6·····
6·····
7·····
8·····友達と 行く·····
9·····
10·····料理は 期待はずれ·····
10·····
10·····
10·····パンフレットの 写真の 料理·····
11·····パンフレットの 写真の お料理·····
11·····食事で 病気に なる·····
12·····お料理で 食中毒を 起こす·····
12·····新鮮な 海の幸·····
13·····
14·····写真·····
15·····大きな 旅館·····新しい 旅館·····手入れの 行き届いた 宿·····
16·····従業員の 態度·····
17·····
18·····
19·····
20·····
20·····
21·····
22·····
23·····
24·····
25·····
26·····
27·····
28·····
29·····
30·····
31·····

A-8

special fragment:

10----- ---- ---------- ---------- ---------- --- ---------- -------
10--- ----- --- -- ----- --- ---------- ---- --------- ---- ----- -----
10 ------ -- ---------- ------ ------ ---------- -- ---------- --- --
11---- -- --- ------ ----- --- ---------- ---- --------- -- ----- ---
11You didn't get sick because of the food, did you?-- ------- -------
12No, I didn't get poisoned by the foods, -- ---- ----- --- ----------
12----- ---- ------ ----- --- ---------- ---- --------- ---- ---- ----
13-------- ----- ------- --- -------- -- -- ---------- -------- -------- -
14---- -- ---- ------ ----- --- --------- --- ---- --------- ---- ----- ---
15----- ---- ------ ----- --- ---------- ---- --------- ---- ----- ----

special fragment:

10·····
10·····
10·····
11·····
11食事のせいでご病気になられたとかそういうことではないんですね。
12ええ、別にお料理で食中毒を起こしたわけじゃありません。·····
12·····
13·····
14·····
15·····