

A Comprehensive and Practical Model of Memory-Based Machine Translation*

Hiroaki Kitano

Center for Machine Translation
Carnegie Mellon University
Pittsburgh, PA 15213 U.S.A.

Abstract

This paper presents a comprehensive and practical model of memory-based machine translation (MBMT), also called example-based machine translation (EBMT). Although MBMT is gaining increasing interest, models proposed so far are incomplete in that these models focus on only one aspect of translation, and ignore involvement of rules which are available as expert knowledge. Goals of this paper are to present (1) a comprehensive architecture for MBMT to serve as a basis for discussion on how real systems should be built, (2) a novel method for integrating memory-based approach and rule-based approach, which would enable handling linguistic phenomena, such as zero-pronoun, and use of expert knowledge available as rules, (3) an optimal representation of translation examples, so that users can incrementally customize the system, (4) a robust matching scheme, using a version of Dynamic Programming, to attain robust and broad-coverage translation. A part of the model has been implemented on massively parallel machines. At the same time, the preliminary evaluation confirmed the improvements on the translation quality and user acceptance over pure MBMTs.

1 Introduction

This paper presents a comprehensive and practical model of Memory-Based Machine Translation (MBMT). The proposed model (1) integrates memory-based process and rule-based process in a novel fashion, (2) employs a user customizable example representation, and (3) introduces a robust and dynamic matching of examples against the input sentence. Although there are several scientific and engineering claims in the proposed model, this paper explicitly focuses on the issue of how an overall system should be organized, rather than discussion on specific issues. This is based on the authors' recognition that the presentation of the comprehensive model is of prime importance at this moment, in order

*This research has been supported by National Science Foundation grant MIP-90/09111, Pittsburgh Supercomputing Center IRI-910002P, and was carried out in conjunction with the International Consortium for Massively Parallel Advanced Computing Technologies (IMPACT).

System	Task	Accuracy
MBT-I [Sato, 1991b]	Word Choice	85.9%
MBT-II [Sato and Nagao, 1990]	Transfer	N.A.
EBMT [Sumita and Iida, 1991]	Phrase (A の B)	89%
TDMT [Furuse and Iida, 1992]	Sentence	88.9%

Table 1: Major MBMTs and their accuracy

to bring MBMT research into the next phase, which involves the development of field-testable systems.

MBMT, also called Example-Based Machine Translation (EBMT) or Case-Based Machine Translation (CBMT), is the state-of-the-art technology enabled by high performance computing and large data resources. The idea was first proposed by Nagao as *Translation by Analogy* [Nagao, 1985]. Nagao's idea parallels memory-based reasoning [Stanfill and Waltz, 1986], case-based reasoning [Reisbeck and Schank, 1990], and derivational analogy [Carbonell, 1986], since the underlying thrust of the model is to use examples of past translation to produce translation of the given input through an analogical process. Take an example of English-Japanese translation, where an English sentence is translated into Japanese through the following process. First, the system retrieves one or more English sentences in the example-base, which are similar to the input sentence. Next, it identifies a combination of examples, which covers the input. Then, adaptation operations will be derived. Finally, it reconstructs Japanese translation by conducting recombination and adaptation on the Japanese part of the retrieved examples.

MBMT counters the traditional machine translation approaches, which rely on grammar rules. However, emergent recognition on the problems regarding the traditional approach (such as tractability, scalability, and performance) and some early successes in high accuracy translation (Table 1) and high performance implementations ([Kitano and Higuchi, 1991a, Kitano and Higuchi, 1991b, Kitano, 1991]) lead the MT community to seriously consider using MBMT as a viable option for next generation MT systems.

Nevertheless, the practicality of MBMT is open to question. Models proposed so far and systems based on these models only dealt with a part of the translation process as seen in Table 1. TDMT (Transfer-Driven Machine Translation [Furuse and Iida, 1992]) is the only system which handles the entire translation process. However, TDMT suffers from coverage requirements, since

no mechanism has been provided to recombine fragments of examples. Plus, none of these models address issues of user customization, robustness, contextual processing, and integration of example-based process and explicit rules for translation.

As stated at the outset of this paper, the goal of this paper is to present a first comprehensive and practical model which addresses these issues, which thereby illustrates how MBMT can be developed as a self-contained MT system. Since the author believe that the presentation of the whole architecture should be given the highest priority, this paper exclusively focuses on overall architecture and how each subcomponent functions. Implementation details and evaluation will be described in a series of papers to be published elsewhere.

2 Overall Architecture

The architecture is shown in Figure 1. The major differences from other MBMTs are the monitoring process and the grammatical inference process. Pure MBMT advocates an extreme-end of viewing translation as a sole memory-based process. However, it is widely recognized that rules are obviously involved in the translation process. the authors' model makes use of rules and linguistic knowledge in two ways.

First, the grammatical inference module uses a local and relation-driven control to infer head, subject, object, and focus of the sentence. The result of inference is accessible from the adaptive translation module and the monitor module. The adaptive translation module uses information on what the *Center* of the previous sentence is, in order to handle zero-pronoun in Japanese-English translation. Identification of the center will be based on the *centering constraint* [Kameyama, 1988]. The monitor module uses information on what is the subject or object of a certain verb in order to decide which expert rules to apply.

Second, the monitor module uses a set of rules encoding translator's knowledge to ensure correct word choice and stylistics. The monitor module initially checks to determine if any word in the sentence matches any part of the condition part of the rules. If there are rules which involve words in the sentence, the monitor module dispatches a request to the grammar inference module, to check if the condition part of the rule (such as *object of increase is short-term interest rate*) can be met. The grammatical inference module uses its local rules and relation-driven control to check whether or not this condition can be satisfied. Then, the grammar inference module returns the result. Depending on the query result, the monitor module may invoke its rules to rewrite the translated sentences.

Although the model uses the rule-based process, this process carries out very different tasks from those employed in the traditional rule-based MT. In fact, there is no place in the proposed model where a complete parse tree is to be built or a full parse is to be carried out. This approach coincides with recent research in MUC-3 [Sundheim, 1991], MUC-4, and TIPSTER, which heavily rely on partial parse and dynamic control strategy, such as relation-driven control [Jacobs, 1992].

3 Knowledge Sources

3.1 Memory-Base

A memory-base consists of examples of translation pair. A translation pair (TX) consists of a source sentence ex-

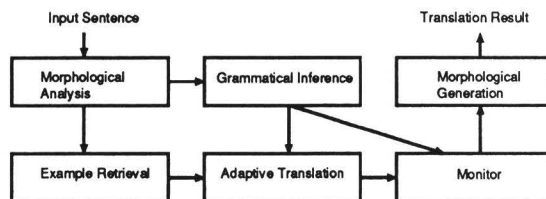


Figure 1: Overall Architecture

ample (SSE), its translation (target sentence example: TSE), and a segment map, which defines segment-level correspondence between SSE and TSE. In the real system, sentences are stored after morphological analysis.

For example, translation examples, seen in Table 2, are stored in the memory-base after the morphological analysis (Table 3).

The segment map defines which SSE segment corresponds to which TSE segment. In case of TX-3, I in SSE has a segment position 1. Since the segment map defines 1 = 1, I corresponds to '私' in TSE, which has segment position 1. The term 'give up' is an idiomatic expression, which should be treated as one word. The segment map defines that 'give up' corresponds to 'あきらめる' by notation 2-3 = 3. 2-3 = 3 denotes that SSE segments from 2 to 3 correspond to segment 3 in TSE.

The model uses segment mapping at the surface level, rather than at the parse tree or canonical representation, as seen in current MBMT models. There are three reasons why the author chose to define mapping at this level, rather than at parse tree or at canonical representation. First, most non-linguists cannot understand what is represented in the parse tree or the canonical representation. Although professional translators are experts in translation, they are not necessarily experts in linguistics. Marking the correspondence at the surface level, however, is possible by many potential users and was reported as being preferred by users in interviews with the author. Second, deriving the correct parse tree (for both SEE and TSE) requires extensive work. Automating this process is not possible at the current level of technology. Even if a large tree-bank were available for translation pairs, user customization will not be possible, as users are unlikely to understand and work on this representation. Third, use of parse tree or canonical representation as a starting point of the memory-based process leaves parsing problems unresolved. As is well understood among commercial MT systems developers, ambiguities and difficulties in tracking the systems behavior during parsing are the major bottlenecks in the quality improvement. Any MBMT approach, which would leave this process on the traditional model, would not attain significant improvement over traditional MT systems.

3.2 Abstraction Hierarchy

There are several ways to calculate the distance between words. A methods to use an abstraction hierarchy, or Thesaurus, (left side of Figure 2) is readily available [Sumita and Iida, 1991, Sato, 1991a]. The problem of using a hand-crafted abstraction hierarchy is that it is difficult to fine-tune the distance between words, which will reflect the real corpus. Alternatives are the statistical or neural network method [McLean, 1992]. However, these

1	I gave a talk.	私は、講演をした。
2	I gave her a gift.	私は、彼女に、プレゼントを渡した。
3	I gave up.	私は、あきらめた。
4	I gave her a lesson.	私は、彼女に教訓を与えた。
5	He gave an example.	彼は、例を挙げた。
6	I am not sure if it was snowing.	雪が降っていたかどうか、良く分からない。
7	Shall I send you a document	資料を、送みましょうか。
8	I already have a document	もう持っています。
9	ODR was increased	公定歩合が引き上げられた。
10	The increase in CD-rate	C Dレートの上昇。

Table 2: Examples of Translation Pair

English Sentences								
No.	1	2	3	4	5	6	7	8
1	I	give -PAST	a	talk				
2	I	give -PAST	her	a	gift			
3	I	give -PAST	up					
4	I	give -PAST	her	a	lesson			
5	He	give -PAST	an	example				
6	I	am	not	sure	if	it	be -PAST	snowing
7	Shall	I	send	you	a	document		
8	I	already	have					
9	ODR	be -PAST	increase -PAST					
10	The	increase	in	CD-rate				

Japanese Sentences								
No.	1	2	3	4	5	6	7	8
1	私	は	講演		する -PAST			
2	私	は	彼女	に	プレゼント	を	渡す -PAST	
3	私	は	あきらめる -PAST					
4	私	は	彼女	に	教訓	を	与える -PAST	
5	彼	は	例	を	挙げる -PAST			
6	雪	が	降っている -PAST	かどうか	良く分からない			
7	資料	を	送る -GER	ましょうか				
8	もう	持っています						
9	公定歩合	が	引き上げ -PSV -PAST					
10	C Dレート	の	上昇					

Segment Map								
No.	1	2	3	4	5	6	7	8
1	1=1	2=5	4=3					
2	1=1	2=7	3=3	5=5				
3	1=1	2-3=3						
4	1=1	2=7	3=3	5=5				
5	1=1	2=5	4=3					
6	1-5=4-5	6-8=1-3						
7	2=0	3=3	4=0	6=1				
8	1=0	2=1	3=2	5=0				
9	1=1	2-3=3						
10	2=3	4=1						

Table 3: A Part of Memory-Base

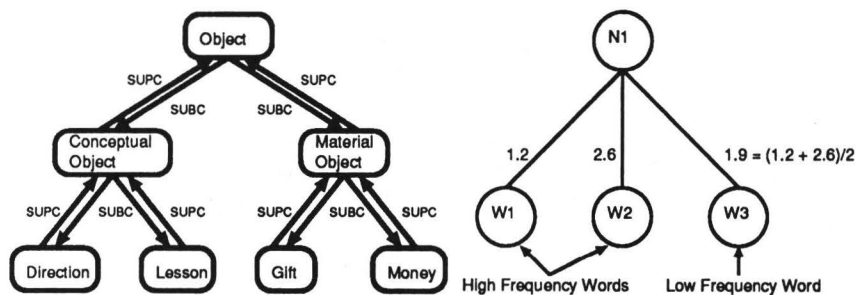


Figure 2: Abstraction-based Word Distance Definition

methods cannot provide accurate distance measures for low-frequency words. For example, these methods cannot determine the appropriate distance for words which appear only once in the corpus. Also, methods using neural networks involve problems in re-training, when new words and examples are added.

The approach used in the MBMT is knowledge-based clustering (KBC). KBC uses a hand-crafted hierarchy or network as a basis of distance calculation, and uses a corpus-based similarity clustering to determine the distance to be assigned to each arc. For low frequency words, a distance is set to the average of distance of other words under the same node. For example, if W1 and W2, both under the same node, have equal distances to nodes 1.2 and 2.6, respectively, a low frequency word W3 under the same node will have 1.9 distance to the node (right side of Figure 2).

4 Grammatical Inference

The grammatical inference module takes an input from the morphological analysis module, which contains basic linguistic information word-stem, tense, part-of-speech, etc. For example, a sentence 'CD-rate was increased' is represented as ((cd-rate () N) (be (-PAST) V) (increase (-PAST) V)) when it is given to the grammatical inference module. When there are multiple possibilities in the part-of-speech and other information, all candidates will be listed. The grammatical inference module uses a local and relation-driven control to identify the head and the fillers for certain relations. For example, when the request is dispatched from the monitor module to check if the object of increase is CD-rate, the grammatical inference module first spots two words 'increase' and 'cd-rate'. If these words cannot be found, nil will be returned. After confirming the existence of these words, local grammar rules are invoked to check whether or not the requested relation can be confirmed. In this case, the relation can be confirmed, because 'increase' is in -PAST form and it is preceded by (be (-PAST) V) which, in turn, was preceded by (cd-rate () N). There are several ways to encode linguistic knowledge without creating a full parse of the sentence. However, the best and theoretically well established method is a subject for further study. At present, the author uses a combination of an island-driven parser and ad-hoc hand-coded rules.

5 Examples Retrieval

Identification of similar examples plays an important role determining the success of MBMT. Assuming that there is no alignment disposition between two sentences, the distance between S_i and S_j is given by:

$$d(S_i, S_j) = \sum_{p=1}^n d(w_{pi}, w_{pj}) \quad (1)$$

where n is the length of the sentence. $d(w_{pi}, w_{pj})$ is a distance between words w_{pi} and w_{pj} . Similarity of the environment, in which the sentence fragment is embedded, should also be a factor. Issues on what makes the best environment similarity measure are presently under experiment, thus they are discarded from the description in this paper.

Assume that an input is: 'You gave me a direction.' TX-2 and TX-4 results in the best match (Table 4).

	Distance	Sentence
Input	0.0	You gave me a direction
TX-4	0.9	I gave her a lesson
TX-2	1.4	I gave her a gift

Table 4: Examples matched for a simple input

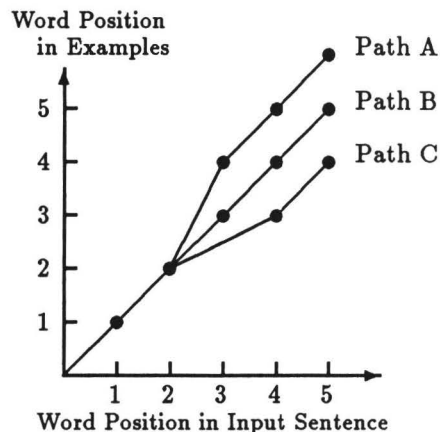


Figure 3: DP-Matching of Input and Examples

This is the simplest case, where there is no alignment dislocation and a single example covers the input sentence.

Matching the examples against the input sentence is based on a version of parallel DP-matching. Figure 3 shows the basic idea. Each word in the input sentence is matched against all examples in the memory-base. When a word matches (an exact match or similarity match) the next word in the input is matched against the next word in the candidate examples. In Figure 3, words in positions 1 and 2 in the input matched with words in positions 1 and 2 in examples. However, a word in position 3 in the input only matched with example B. Example A has an extra word inserted in between words in positions 2 and 3 of the input, and example C misses a word in position 3 of the input. For example, let us assume that example A is 'a b f c d e', example B is 'a b c d e', example C is 'a b d e', and the input is 'a b c d e'. Example B results in a complete match, which is shown as a straight line in Figure 3. Example A has an extra word 'f' in between 'b' and 'c'. Thus, 'c' in the input, which is at position 3 (the horizontal axis), matches with 'c' in the example, which is at position 4 (the vertical axis). Since this matching is not exact, it incurs some cost. However, it is still considered as a candidate. The same is true for example C. Due to space limitations, it is not feasible to describe the details of this matching mechanism. Interested readers should consult literature on the DP-Matching and Dynamic Time Warping for basic ideas and mathematics. Multiple matching will take place in real sentences, as seen in Figure 4.

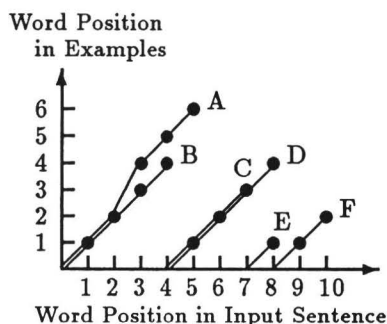


Figure 4: Multiple Match between Examples

Difference Table	
TX-4	Input
I	You
her	me
lesson	direction

Table 5: Difference Table

6 Adaptive Translation

Once the system has retrieved similar cases, the next step is to synthesize a derivation from the input sentence to the translation. This process is divided into (1) covering the input sentence by retrieved cases, (2) producing an adaptation operation list, and (3) reconstructing the translation from the target language part of the retrieved cases.

6.1 Adaptation with a Single Case

This process can be illustrated by using the previous example. For the input sentence, two examples (TX-4 and TX-2) are retrieved. Since a single case can cover the input sentence, the system simply uses TX-4, which is the most similar case which covers the input sentence. An example of using multiple cases to cover an input sentence will be shown later.

Next, the difference between the input sentence and TX-4 is checked. Table 5 shows the differences between the input sentence and TX-4. In this case, the input sentence and TX-4 are sufficiently similar so that only a lexical level adaptation would be sufficient to produce a translation.

Then, adaptation operations are derived, based on the difference table. For example, a word 'I' in TX-4 needs to be replaced with 'you' to cover the input sentence. This operation is associated with change of 私 ('I') into あなた ('you') in a target language part of TX-4. Table 6 shows adaptations involved in this translation. Word choices may be based on a statistical likelihood, computed from

Adaption Operations			
English		Japanese	
I	→ You	私	→ あなた
her	→ me	彼女	→ 私
gift	→ direction	教訓	→ 行き方

Table 6: Adaptation Operations

segment maps. In effect, the method is similar to MBT-I [Sato, 1991b].

Once the adaptation operations are defined, the final stage is to reconstruct translation from the target language part of the cases. In this example, TX-4 is the only case involved. Thus, TSE for TX-4 will be adapted, using the derived adaptation operations. This process is shown in Table 7¹.

6.2 Combination of Cases

For more complex sentences, the input sentence needs to be covered by multiple examples. Figure 4 shows an example of a matching diagram for such an event. Examples from A to F provide partial matches to the input, and several combinations to cover the input are possible. The proposed model chooses to use the minimum cost covering method which takes into account size, similarity, and DP-matching cost of each fragment. Heuristics for scoring the combination of examples and an algorithm for the minimum cost covering for tree structured examples has been proposed [Sato and Nagao, 1990, Maruyama and Watanabe, 1992]. The proposed model employs the 1-D version of these algorithms. In the given example, a combination of examples B D F is likely to be preferred over a combination of B C E F, A C E F, etc. This is because B D F covers the input with a minimum number of examples and no example involves the disposition in DP-Matching.

Assume we have an input: 'I am not sure if she gave him a gift'. None of the cases in the memory-base cover the input sentence; even the best match cases only covers part of the input sentence. Table 8 shows the input sentence and retrieved cases. TX-6 matches only in sentence segments from 1 to 5, 'I am not sure if', and TX-2 matches from 6 to 10, 'she gave him a gift'.

At this moment, which segments can be used in each case can be defined using the segment map. In this case, TX-1 has segment of 1-5 ('I am not sure if') and 6-8 ('it was snowing'). Therefore, an obvious operation is to use 'I am not sure if' and 'she gave him a gift'. Since a part of input sentence is 'I gave her a gift', 'she gave him a gift' has to be adapted. The segment map is used again, to identify which words should be replaced. As a result of these operations, adaptation operations are defined, and this is cross applied to produce Japanese translation.

Since segment 1 of the TX-1 has high similarity and segment 2 has low similarity, only segment 1 will be used, and the TX-2 will be used to translate 'I gave her a gift.' Table 9 shows how adaptive translation is carried out in this example.

6.3 Zero Pronoun Handling

Substantial percentage of Japanese sentences involves zero-pronoun (a kind of ellipsis) which need to be resolved and articulated in English translation. The following dialogue illustrates the problem of zero-pronoun (a word in [...] was not articulated.):

Speaker-1: 資料を送りましょうか (Shall [I] send [you] a document)

Speaker-2: もう、持っています ([I] already have [a document].)

¹For the sake of simplicity, the morphological part is ignored in this example.

	1	2	3	4	5	6	7
Best Match Example	私	は、	彼女	に	教訓	を	与えた。
Adaptation	私 → あなた		彼女 → 私		教訓 → 行き方		
Translation Result	あなた	は、	私	に	行き方	を	与えた。

Table 7: Adaptation for a simple sentence translation

	1	2	3	4	5	6	7	8	9	10
Input	I	am	not	sure	if	she	gave	him	a	gift
TX-6	I	am	not	sure	if	it	was	snowing		
TX-2						I	gave	her	a	gift

Table 8: Retrieved Examples

The authors' solution to this problem is to explicitly define the zero-pronoun in the segment map. TX-7 and TX-8 in Table 3 show translation pairs for sentences with zero-pronoun. Words which are not articulated in Japanese are marked as 0 in the segment map. In the Japanese to English translation process, a segment whose Japanese correspondence is 0 needs to be recovered using a list of the center. Information on what is in the center is provided from the grammatical inference module.

7 Monitor

The monitor process checks whether or not the translation result is stylistically and grammatically appropriate. This process has not been incorporated in previous models of MBMT. The main rationale for the monitor process is the existence of the domain rules for translation. These rules are widely used by professional translators and are explicitly documented. For example, in the economic domain, an English word 'prediction' must be translated as 予想 when no specific number is given, but the same word must be translated as 予測 when specific numbers are given. Also, the word 'increase' must be translated as 引き上げ (pull up) in the context of 'increase in the official discount rates', but it should be 上昇 (elevation) for 'increase in the short-term interest rates'. Thus, 短期金利の引き上げ (pull up in short-term interest rates) is considered as being a mistranslation. An extreme example is wherein the use of the word 投資銀行 (investment bank) for Solomon Brothers and Goldman-Sacks is officially unacceptable for the Japanese government as these banks are approved as security companies, not as investment banks. At this time, the English word 'investment bank' must be phonetically transcribed as インベストメント・バンク.

Segregating these rules from examples is extremely inefficient, even if it were possible. Given the fact that professional translators, who are the users of the system, have these rules explicitly written, the most efficient and user acceptable way is to directly use these rules in the translation process. The main task of the monitor process is the monitoring of translation, using these domain-specific rules. For example, a rule specifies that, if there is a word 'increase' in the input and the object of 'increase' is 'CD-rate', then 'increase' should be translated as '上昇', an internal representation (not exposed to users) is shown below:

((COND (AND (EQUAL HEAD 'increase'))

(EQUAL OBJECT 'CD-rate'))))
(ACTION (TRANSLATE 'increase' '上昇'))))

Using the examples in Table 3, conventional MBMT systems would translate 'CD-rate was increased' into 'CDレートが引き上げられた.' Although TX-10 implicitly encodes knowledge that 上昇 is associated with CDレート, it did not contribute to this translation because its did not match with the input sentence due to the substantial difference in their syntactic structures. Since CD-rate is a kind of short-term interest rate, this translation is incorrect. In the model, the output from the adaptive translation module will be checked against expert rules. The monitor module search words appear in rules in the input sentence. If there is any rule which involves words in the sentences, requests are dispatched to the grammatical inference module to verify whether or not the condition to invoke the rule can be met. In this case, the condition is met (as explained in Section 4), so that the word 引き上げられた will be replaced with 上昇した. Thus, the final translation will be CDレートが上昇した which is a correct translation. Currently, there are 82 of such rules for the financial domain.

8 Preliminary Evaluation

A brief report on the initial evaluation phase would help understanding superiorities of the model. Details will be reported elsewhere. The author have implemented an experimental prototype of English-Japanese translation for the financial domain. Example matching and retrieval module have been implemented on Semantic Network Array Processor (SNAP) [Kitano et. al., 1991a] and the CM-2 connection machine. Unoptimized version resulted in few hundreds milli seconds order performance. The performance figure is almost constant regardless of the memory-base size, due to data-parallel nature of the algorithm. While a major part of the computing time has been spent on interprocessor communication, optimization would certainly provide better performance and scalability. Independently, joint research with ATR and the Electrotechnical Laboratory obtained promising performance for phrase translation tasks [Sumita et. al., 1993].

Independently, the effects of the monitor process were examined, using 100 sentences from the Wall Street Journal (WST) and The Financial Times (FT). 95 words, which are considered to be problematic in translation, have been selected as a benchmark. While semantic accuracy was 100%, there are stylistic problems, as discussed in Section 7. Pure MBMT attained stylistically

	1	2	3	4	5	6	7	8	9
TX-6 TX-2	私	は	彼女	に	プレゼント	を	あげた	かどうか	良く分からない
Adaptation-1	雪が降っていた → 私は彼女にプレゼントをあげた								
Intermediate	私	は	彼女	に	プレゼント	を	あげた	かどうか	良く分からない
Adaptation-2	私 → 彼女 彼女 → 彼								
Translation	彼女	は	彼	に	プレゼント	を	あげた	かどうか	良く分からない

Table 9: Adaptive Translation Process

correct translation for 48 words out of 95 words. When it is combined with the expert rules (using the monitor module), 84 words out of 95 words were translated with correct stylistics. The rule in the monitor module has been invoked in 76 cases. Some 40 cases were consistent with the pure MBMT translation, while 36 cases resulted in modification of translation. 9 words were translated with incorrect style due to lack of examples and rules. This figure, however, should not be viewed as indicating the MBMT accuracy, because it focuses on problematic events. It should be viewed as evidence to support the effectiveness of the monitor process, and should not be used to draw any quantitative conclusions. With larger memory-base and more detailed rules, the accuracy is expected to improve drastically.

9 Conclusion

In this paper, the author proposed a comprehensive and practical model of memory-based machine translation. The major contributions in this paper are: presentation of the first comprehensive and practical model of MBMT, and introduction of the novel architecture to integrate memory-based and rule-based processes.

In effect, the model offers a basis for achieving a solution towards user customizable example definition, handling contextual phenomena, such as zero-pronoun, broad-coverage and robust translation, and use of expert knowledge.

Use of expert knowledge in the rule form is the key factor which mitigated the problem of sparse corpus inherent in the early phase of deployment. Although a more formal model needs to be developed on how local grammar and expert knowledge should be encoded, preliminary evaluation suggests the monitor process actually improves the quality of translation.

Massively parallel implementation of the example retrieval and matching attained promising performance. This implies that MBMT can attain both high speed and high quality translation.

The authors hope the model presented here will serve as a basis for further discussion on how MBMT should be developed as a self-contained system, which is expected to be a mainstay approach in the next generation machine translation.

References

- [Carbonell, 1986] Carbonell, J. G., "Derivational Analogy: A theory of Reconstructive Problem Solving and Expertise Acquisition," in *Machine Learning, An Artificial Intelligence Approach, Volume II*, Michalski, R. S., Carbonell, J. G., and Mitchell, T. M., eds., Morgan Kaufmann, 1986.
- [Furuse and Iida, 1992] Furuse, O. and Iida, H., "Cooperation between Transfer and Analysis in Example-Based Framework," *Proc. of COLING-92*, 1992.
- [Kameyama, 1988] Kameyama, M., "A Property-Sharing Constraint in Centering," *ACL-88*, 1988.

- [Kitano, 1991] Kitano, H., "ΦDm-Dialog: an Experimental Speech-to-speech Translation System," *IEEE Computer*, June, 1991.
- [Kitano and Higuchi, 1991a] Kitano, H. and Higuchi, T., "High Performance Memory-Based Translation on IXM2 Massively Parallel Associative Memory Processor," *Proc. of AAAI-91*, 1991.
- [Kitano and Higuchi, 1991b] Kitano, H. and Higuchi, T., "Massively Parallel Memory-Based Parsing," *Proc. of IJCAI-91*, Sydney, 1991.
- [Kitano et. al., 1991a] Kitano, H., Moldovan, D., and Cha, S., "High Performance Natural Language Processing on Semantic Network Array Processor," *Proc. of IJCAI-91*, Sydney, 1991.
- [Maruyama and Watanabe, 1992] Maruyama, H. and Watanabe, H., "Tree Cover Search Algorithm for Example-Based Translation," *Proceedings of TMI-92*, 1992.
- [Jacobs, 1992] Jacobs, P., "Parsing Run Amok: Relation-Driven Control for Text Analysis," *Proc. of AAAI-92*, 1992.
- [McLean, 1992] McLean, I., "Example-Based Machine Translation using Connectionist Matching," *Proceedings of TMI-92*, 1992.
- [Nagao, 1985] Nagao, M., "A Framework of a Mechanical Translation between Japanese and English by Analogy Principle," 1985.
- [Reisbeck and Schank, 1990] Reisbeck, C. and Schank, R., *Inside Case-Based Reasoning*, Lawrence Erlbaum Associates, 1990.
- [Sato, 1991a] Sato, S., *Example-Based Machine Translation*, Ph.D. Thesis, Kyoto University, 1991.
- [Sato, 1991b] Sato, S., "MBT-I: Word Choice based on Examples," *Journal of Japanese Society for AI*, Vol. 6, No. 4, 1991 (in Japanese).
- [Sato and Nagao, 1990] Sato, S. and Nagao, M., "Toward Memory-Based Translation," *Proc. of COLING-90*, 1990.
- [Sumita and Iida, 1991] Sumita, E. and Iida, H., "Experiments and Prospects of Example-Based Machine Translation," *Proceedings of the 29th annual Meeting of the Association for Computational Linguistics*, 1991.
- [Sumita et. al., 1993] Sumita, E. Oi, K., Iida, H., Higuchi, T., Takahashi, N., Kitano, H., "Example-Based Machine Translation on Massively Parallel Processors," *Proc. of IJCAI-93*, 1993.
- [Sundheim, 1991] Sundheim, B. (Ed.), *Proceedings of the Third Message Understanding Conference (MUC-3)*, Morgan Kaufmann Publishers, San Mateo, CA, 1991.
- [Stanfill and Waltz, 1986] Stanfill, C. and Waltz, D., "Toward Memory-Based Reasoning," *Communications of the ACM*, Vol. 29, No. 12, 1986.