An Analysis of Translation Equivalents of Japanese-Mongolian Functional Expressions

Sanduijav Enkhbayar and **Takehito Utsuro** and **Satoshi Sato**
Graduate School of Informatics, Kyoto University
Sakyo-ku, Kyoto, 606-8501
{enkh,utsuro,sato}@pine.kuee.kyoto-u.ac.jp

Abstract

This paper studies an issue of translating Japanese functional expressions into Mongolian, as a step toward building a Japanese-Mongolian MT system. We first examine whether each Japanese functional expression can be translated into a Mongolian functional expression unambiguously without considering the context of the Japanese expression. The rate of unambiguous translation is quite high (86 % in total). We also show that the Japanese and Mongolian languages are similar in usages as well as in syntactic construction of functional expressions. Finally, we show that, even with simple one to one table lookup translation rules without considering contexts, about 70 $\sim$ 90 % of Japanese functional expressions can be translated into Mongolian with the accuracy rate nearly 95%.

1 Introduction

This paper studies an issue of translating Japanese functional expressions into Mongolian, as a step toward building a Japanese-Mongolian MT system. Japanese and Mongolian are both head final, SOV languages and the sentence structures of the two languages are quite similar. Between such a language pair, word orders are usually preserved after translation, so it is expected that syntactic ambiguities of a source sentence are not required to be resolved in translation. In such a language pair, major difficulties in MT arise in target word selection, both for content words and for functional expressions. Considering this situation, we focus on translation equivalents of Japanese-Mongolian functional expressions, which are more complicated than translation equivalents of content words, and thus could be one of the most important issues when building an MT system of this language pair.

As a first step, we examine Japanese functional expressions listed in the test content specifications of the Japanese language proficiency test for Japanese learners (Association of International Education, Japan, 2002). The Japanese language proficiency test is graded in four levels (from level 4 : lowest to level 1 : highest). For each level, the test content specification lists about 100 $\sim$ 200 functional expressions, out of which we examine those of levels 2, 3, and 4. We examine whether each functional expression can be translated into a Mongolian functional expression unambiguously without considering the context of the Japanese expression. The rate of unambiguous translation is 86 % in total, and thus is quite high.

For each of the levels 2, 3, and 4, we randomly pick up 30 Japanese functional expressions from those which can be unambiguously translated into Mongolian without considering the context, and evaluate the accuracy of translating those 30 functional expressions in their example sentences (620 sentences in total, taken from (Group Jamashii, 1998)). The rate of correct translation is 80%, and that of (partially) correct translation is 95%.

This paper presents the details of the analysis above and discusses further issues toward real-

(1) noun stem
хүүхэд (*child*)
(2) noun stem + suffix (number)
хүүхдүүд (*children*)
(3) noun stem + suffix (number) + case marker
хүүхдүүдэд (*to children*)
(4) noun stem + suffix (number) + case marker + reflexive possessive
хүүхдүүддээ (*to one's children*)

Figure 1: Examples of Mongolian Noun Stem + Suffixes

(1) verb stem
ид (*eat*)
(2) verb stem + passive voice
идэгд (*be eaten*)
(3) verb stem + causative voice
идүүл (*let someone eat*)
(4) verb stem + suffix (intention)
идье (*let's eat*)
(5) verb stem + suffix (past)
идэв (*ate*)
(6) verb stem + suffix (perfect)
идчихсэн (*have eaten*)
(7) verb stem + suffix (subordinate clause)
идтэл (*until one eats*)
(8) verb stem + passive voice + suffix (past)
идэгдэв (*was eaten*)

Figure 2: Examples of Mongolian Verb Stem + Suffixes

Table 1: POS Sequence Patterns of Mongolian Functional Expressions

POS sequence patterns	# of functional expressions
suffix + auxiliary verb	45 (14.7%)
auxiliary verb	32 (10.4%)
adverb	23 (7.5%)
suffix	19 (6.2%)
suffix + formal noun + auxiliary verb	13 (4.2%)
suffix + post-positional particle	13 (4.2%)
formal noun	9 (2.9%)
post-positional particle	8 (2.6%)
formal noun + adverb	8 (2.6%)
interrogative	8 (2.6%)
others	129 (42.1%)
total	307 (100%)

Table 2: Ambiguities in Translating Japanese Functional Expressions into Mongolian

level	no target expression	ambiguous	unambiguous	total
2	0 (0.0%)	12 (5.9%)	193 (94.1%)	205 (100%)
3	2 (1.7%)	14 (12.0%)	101 (86.3%)	117 (100%)
4	8 (6.0%)	28 (21.1%)	97 (72.9%)	133 (100%)
total	10 (2.2%)	54 (11.9%)	391 (85.9%)	455 (100%)

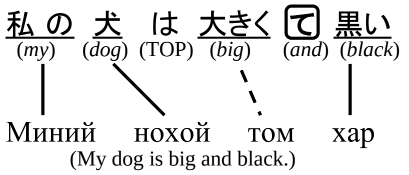
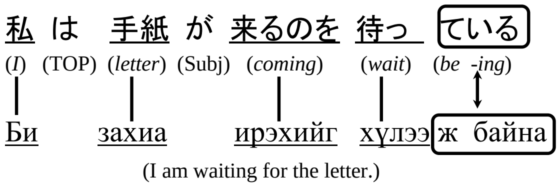

私 の 犬 は 大きく て 黒い
(my) (dog) (TOP) (big) (and) (black)
Миний нохой том хар
(My dog is big and black.)

Figure 3: An Example of “no target expression” (functional expression = “て” (*and*))

a. Continuation


私 は 手紙 が 来るのを 待っ ている
(I) (TOP) (letter) (Subj) (coming) (wait) (be -ing)
Би захиа ирэхийг хүлээж байна
(I am waiting for the letter.)

b. Result

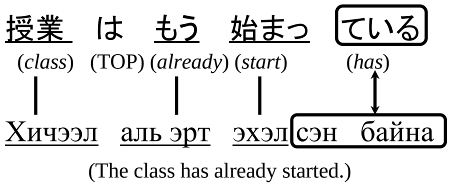

授業 は もう 始まっ ている
(class) (TOP) (already) (start) (has)
Хичээл аль эрт эхэлсэн байна
(The class has already started.)

Figure 4: An Example of “ambiguities in Mongolian translation ” (functional expression = “ている”)

izing Japanese-Mongolian machine translation of functional expressions.

2 Grammar of Mongolian Language

In the Mongolian language, Cyrillic alphabet is used. A sentence consists of a sequence of words, separated from each other by a space. One word consists of a stem of a content word such as a noun and a verb, concatenated with several suffixes. A noun stem is usually followed by a suffix specifying numbers, a case marking suffix, and a reflexive possessive suffix, in this oder. Figure 1 gives several examples. A verb stem is followed by various suffixes such as voice suffixes, imperative suffixes, tense/aspect suffixes, and those introducing subordinate clauses. Figure 2 gives several examples. Word order is generally SOV, and a verb is in the sentence-final position. Word order of constituents other than verbs is relatively free, as in the case of Japanese (Kuribayashi, 1992). As can be seen from the examples of Japanese-Mongolian translation sentences and word alignment in Figures 3 $\sim$ 7, word order of a Mongolian sentence is quite similar to that of its Japanese translation. Table 1 also shows major parts-of-speech sequence patterns of Mongolian functional expressions as well as their distribution.

3 Translation Equivalents of Japanese-Mongolian Functional Expressions

We examine Japanese functional expressions listed in the test content specifications of the Japanese language proficiency test for Japanese learners. For each of the levels 2, 3, and 4 of the Japanese language proficiency test, we examine 100 $\sim$ 200 functional expressions (455 in total).

3.1 Ambiguities in Japanese to Mongolian Translation

We first examine whether each of the Japanese functional expressions can be translated into a Mongolian functional expression unambiguously without considering the context of the Japanese expression. We classify them into the following three categories: 1) no target expression in Mongolian, 2) having ambiguities in target expression selection in Mongolian, 3) no ambiguity in target expression selection in Mongolian. As shown in Table 2, the rates of unambiguous translation are 73% for the level 4, 86% for the level 3, and 94% for the level 2. This result of unambiguous trans-

Table 3: Ambiguity Preserving Rate in "Unambiguous Mongolian to Japanese Translation"

level	preserving ambiguities	unambiguous	total
2	45 (23.3%)	148 (76.7%)	193 (100%)
3	28 (27.7%)	73 (72.3%)	101 (100%)
4	21 (21.6%)	76 (78.4%)	97 (100%)
total	94 (24.0%)	297 (76.0%)	391 (100%)

Table 4: Classification of Literal/Paraphrase in "Unambiguous Mongolian to Japanese Translation"

level	literal translation		paraphrase		total
	$\mathrm{J} = \mathrm{M}$	$\mathrm{J} \subset \mathrm{M}$	$\mathrm{J} = \mathrm{M}$	$\mathrm{J} \subset \mathrm{M}$	
2	106 (54.9%)	22 (11.4%)	54 (28.0%)	11 (5.7%)	193 (100%)
3	79 (78.2%)	14 (13.9%)	6 (5.9%)	2 (2.0%)	101 (100%)
4	65 (67.0%)	23 (23.7%)	8 (8.2%)	1 (1.0%)	97 (100%)
total	250 (64.0%)	59 (15.1%)	68 (17.4%)	14 (3.6%)	391 (100%)

lation rates can be explained as follows: Japanese functional expressions of the level 4 are those normally mastered in the first half of an elementary Japanese language course, some of which have more than one usages. On the other hand, those of the level 2 are normally mastered in the intermediate Japanese language course, most of which are very idiomatic and thus have only one usage.

Figure 3 gives an example of "no target expression in Mongolian", where the Japanese functional expression "て" (*and*) has no Mongolian translation. In Mongolian, adjectives have no conjugation, so only the conjugative form "大きく" (*big*) of the adjective "大きい" (in its base form) is aligned to the Mongolian counterpart. Figure 4 also gives an example of "having ambiguities in target expression selection in Mongolian", where the Japanese functional expression "ている" has two usages (continuation/result, usage distinction is taken from (Group Jamashii, 1998)), and their Mongolian translations are different.

Here, for those Japanese functional expressions which can be unambiguously translated into Mongolian, there could be cases where those expressions actually have multiple usages while the ambiguity can be preserved into Mongolian without resolving it. We classify the ambiguity preserving cases and the unambiguous cases, i.e., the case where the Japanese functional expression has exactly one usage. As shown in Table 3, the ambiguity preserving rate is 16% for the level 4, 32% for the level 3, and 25% for the level 2. Figure 5 gives an example of the ambiguity preserving case, where the Japanese functional expression "てくる" (approaching movement / serial occurrence / appearance / start / action directed towards the viewpoint, usage distinction is taken from (Group Jamashii, 1998)) has five usages while all of them can be translated into one Mongolian expression.

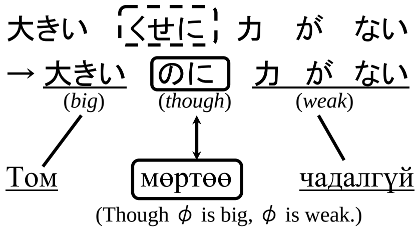


Figure 6: An Example of "Paraphrase / $\mathrm{J} = \mathrm{M}$" (functional expression = "くせに" (*though*))

Next, we classify those Japanese functional expressions which can be unambiguously translated into Mongolian, according to whether each functional expression can be literally translated into Mongolian, or it requires to be paraphrased into

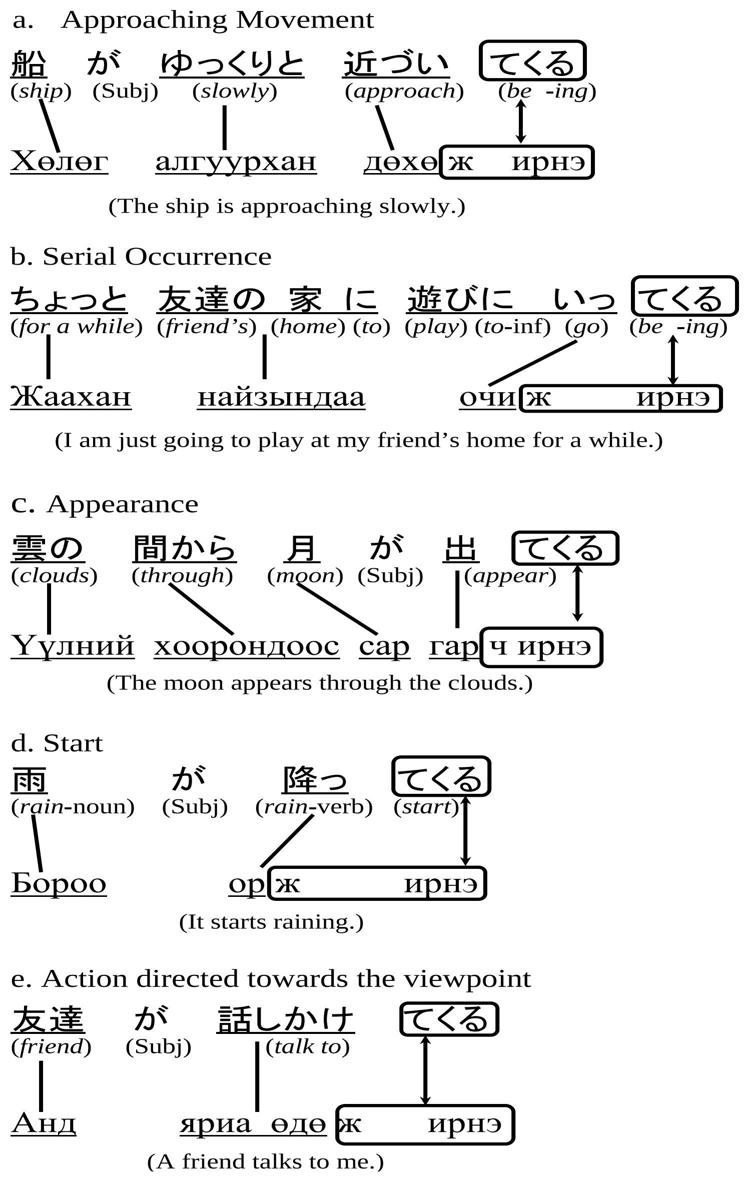

a. Approaching Movement
船 が ゆっくりと 近づい てくる
(ship) (Subj) (slowly) (approach) (be -ing)
Хөлөг алгуурхан дөхө ж ирнэ
(The ship is approaching slowly.)
b. Serial Occurrence
ちょっと 友達の 家 に 遊びに いっ てくる
(for a while) (friend's) (home) (to) (play) (to-inf) (go) (be -ing)
Жаахан найзындаа очи ж ирнэ
(I am just going to play at my friend's home for a while.)
c. Appearance
雲の 間から 月 が 出 てくる
(clouds) (through) (moon) (Subj) (appear)
Үүлний хоорондоос сар гар ч ирнэ
(The moon appears through the clouds.)
d. Start
雨 が 降っ てくる
(rain-noun) (Subj) (rain-verb) (start)
Бороо ор ж ирнэ
(It starts raining.)
e. Action directed towards the viewpoint
友達 が 話しかけ てくる
(friend) (Subj) (talk to)
Анд яриа өдө ж ирнэ
(A friend talks to me.)

Figure 5: Preserving Ambiguities in Mongolian to Japanese Translation (functional expression = “てくる”)

Table 5: Compositionality in “literal translation / J $=$ M”

level	compositional	partially compositional	one word to one word	non-compositional	total total
2	34 (32.1%)	32 (30.2%)	4 (3.8%)	36 (34.0%)	106 (100%)
3	20 (25.3%)	14 (17.7%)	18 (22.8%)	27 (34.2%)	79 (100%)
4	9 (13.8%)	13 (20.0%)	31 (47.7%)	12 (18.5%)	65 (100%)
total	63 (25.2%)	59 (23.6%)	53 (21.2%)	75 (30.0%)	250 (100%)

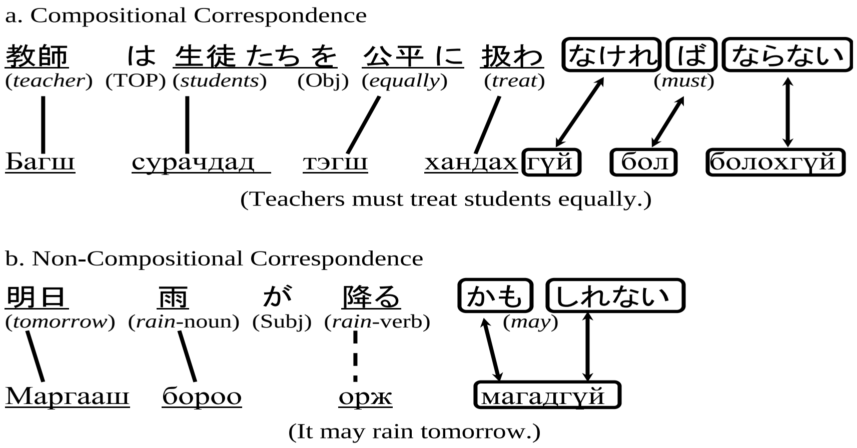

a. Compositional Correspondence
教師 は 生徒 たち を 公平 に 扱わ なけれ ば ならない
(teacher) (TOP) (students) (Obj) (equally) (treat) (must)
Багш сурачдад тэгш хандах гүй бол болохгүй
(Teachers must treat students equally.)
b. Non-Compositional Correspondence
明日 雨 が 降る かも しれない
(tomorrow) (rain-noun) (Subj) (rain-verb) (may)
Маргааш бороо орж магадгүй
(It may rain tomorrow.)

Figure 7: Examples of (Non-)Compositional Correspondences in “literal translation / J $=$ M”

other Japanese functional expression[1] [2]. Then, we further classify each of “literal/paraphrase” categories into two sub-categories: 1) ”J $=$ M”, where the Japanese and the Mongolian functional expressions have one to one mapping, 2) ”J $\subset$ M”, where the Mongolian equivalent has broader usages. Figure 6 gives an example of “paraphrase / J $=$ M”, where the Japanese functional expression of level 2 “くせに” (*though*) can not be literally translated into Mongolian and is paraphrased into “のに” (level 3).

The result of the classification is shown in Table 4. In the “literal translation” category, 64% of the Japanese functional expressions belong to the ”J $=$ M” category. This indicates that, even in the opposite direction of Mongolian to Japanese translation, certain portions Mongolian functional expressions can be unambiguously translated into Japanese without considering context. Rates of those requiring paraphrase are less than 10% for the levels 3 and 4, while for the level 2, the rate is about 34%. However, for the level 2, after paraphrasing them, most of the functional expressions have one to one mapping with Mongolian counterpart, and over 80% of the functional expressions of the level 2 can be regarded as belonging to ”one to one” category.

3.2 Compositional Correspondence of Functional Expressions

For the translation equivalents of Japanese-Mongolian functional expressions of the “literal translation / J $=$ M” category, we examine their compositionality. Here, if both Japanese and Mongolian functional expressions consist of n ($n \leq 2$) words and i-th ($1 \leq i \leq n$) constituent

---

[1] This strategy follows the idea of SANDGLASS MT architecture (Yamamoto, 2002), where, if an expression of the source language is difficult to literally translate into the target language, it is paraphrased into another expression which can be literally translated into the target language.

[2] Here, we restrict the paraphrase as that within the same level of the Japanese language proficiency test, or that into easier levels. Exceptional cases are where a functional expression needs to be paraphrased into an expression that are not listed in the test content specifications of the Japanese language proficiency test. In such cases, we ignore this restriction.

Table 6: Evaluation of Japanese to Mongolian Translation (averaged over 30 functional expressions)

level	correct	partially correct	error
2	89.9%	8.6%	1.5%
3	75.7%	18.9%	5.4%
4	82.4%	11.3%	6.2%
total	82.7%	12.9%	4.4%

Table 7: Evaluation of Japanese to Mongolian Translation (averaged over 620 sentences)

level	correct	partially correct	error	total
2	146 (89.6%)	15 (9.2%)	2 (1.2%)	163 (100%)
3	186 (74.1%)	50 (19.9%)	15 (6.0%)	251 (100%)
4	162 (78.6%)	29 (14.1%)	15 (7.3%)	206 (100%)
total	494 (79.7%)	94 (15.1%)	32 (5.2%)	620 (100%)

words of the two expressions can be literally translated into each other, then the Japanese and Mongolian functional expressions are defined to be compositional translation equivalent. According to the degree of compositionality, translation equivalents in the “literal translation / J = M” category are classified into the following four subcategories: 1) compositional, 2) partially compositional, 3) both Japanese and Mongolian functional expressions consist of one word (one word to one word), 4) non-compositional. Figure 7 gives an example of compositional correspondence and that of non-compositional correspondence.

Table 5 shows the result of this classification. In total, nearly half of the functional expressions equivalents are classified as fully or partially compositional. Furthermore, 23% (level 3) or 48% (level 4) of translation equivalents are pairs consisting of one word functional expression. On the other hand, in the level 2, since most expressions are idiomatic, they consist of more than one words and over 60% of them are compositional translation equivalents. This high compositionality rate indicates that Japanese and Mongolian have quite similar sentence structure even in the composition of functional expressions consisting of multiple words.

4 Evaluation of Japanese to Mongolian Translation

For each of the levels 2, 3, and 4, we randomly pick up 30 Japanese functional expressions from those which can be unambiguously translated into Mongolian without considering the context, and evaluate the accuracy of translating those 30 functional expressions in their example sentences (620 sentences in total, taken from (Group Jamashii, 1998)). These 30 functional expressions are selected so that their distribution follows the distribution of the number of functional expressions in Table 4. Translation accuracy is evaluated in the three levels: 1) correct, 2) partially correct, and 3) error. Results averaged over 30 functional expressions are shown in Table 6, while those averaged over 620 sentences are shown in Table 7.

On the average, fully correct rate is about 80% and (partially) correct rate is about 95%. Translation error rates are about 6 $\sim$ 7 % for the levels 3 and 4, and about 1% for the level 2. Especially, fully correct rate for the level 2 is over 90% and is quite high. These results clearly show that, for Japanese functional expressions which can be unambiguously translated into Mongolian without considering the context, it is quite possible to realize high performance Japanese to Mongolian machine translation of functional expressions.

5 Concluding Remarks

This paper studied an issue of translating Japanese functional expressions into Mongolian, as a step toward building a Japanese-Mongolian MT system. We examined whether each Japanese functional expression can be translated into a Mongolian functional expression unambiguously without considering the context of the Japanese expression. The rate of unambiguous translation is quite high (86 % in total). We also showed that, even with simple one to one table lookup translation rules without considering contexts, about 70 $\sim$ 90 % of Japanese functional expressions can be translated into Mongolian with the accuracy rate nearly 95%. Future plans definitely include the followings: 1) analysis of Mongolian to Japanese translation equivalents of functional expressions, 2) investigating more complicated translation selection rules, i.e., disambiguation rules for Japanese and Mongolian functional expressions, 3) comparative study of translation equivalents of Japanese-English functional expressions.

References

Association of International Education, Japan. 2002. *Japanese Language Proficiency Test: Test Content Specifications (revised edition)*. BONJINSHA Co., Ltd. (in Japanese).

Group Jamashii, editor. 1998. *Nihongo Bunkei Jiten*. Kuroshio Publisher. (in Japanese).

H. Kuribayashi. 1992. Mongolian. In T. Kamei, R. Kawano, and E. Chino, editors, *The Sanseido Encyclopedia of Language, Vol.4, Language of the World，Part 3*, pages 501–517. Sanseido Co., Ltd. (in Japanese).

K. Yamamoto. 2002. Machine translation by interaction between paraphraser. In *Proc. 19th COLING*, pages 1107–1113.