

日本語・モンゴル語間の機能表現の機械翻訳 Japanese-Mongolian Machine Translation of Functional Expressions

Sanduijav ENKHBAVAR

宇津呂武仁

佐藤理史

京都大学大学院 情報学研究科

enkh@pine.kuee.kyoto-u.ac.jp, {utsuro, sato}@i.kyoto-u.ac.jp

1. はじめに

本論文では、日本語・モンゴル語間の機械翻訳を実現するための要素技術の一つとして、日本語・モンゴル語間における機能表現の翻訳の問題に焦点を当てる。モンゴル語¹は日本語と同じ SOV 言語であり、モンゴル語の語順は日本語の語順に極めて近い[亀井 1992]。また、述部が原則として節末に位置し、その他の格要素や副詞など間の語順が比較的自由である点も日本語によく似ている。したがって、日本語とモンゴル語の間の機械翻訳においては、構文解析の曖昧性の問題を考慮せず、単語もしくは句の置き換えによる逐語訳の程度でも、かなりの部分が翻訳できることが期待できる。このような言語間の機械翻訳においては、自立語や機能表現の適切な訳語選択が最も重要な問題となる。

そこで、本論文では、それらの問題のうち、日本語・モンゴル語の双方において文法現象の中心的役割を担う機能表現に注目し、日本語・モンゴル語間における機能表現の翻訳について分析を行う。具体的には、本論文では、日本語の機能表現をモンゴル語に翻訳する方向について、日本語・モンゴル語間での機能表現の訳語対応の曖昧性の度合いを調査した。ここでは特に、日本語能力試験出題基準[日本国際教育協会 2002]の文法項目において掲載されている機能表現をとりあげ、日本語能力試験の 2~4 級の各級に対して、100~200 個の機能表現(合計で 455 個)を対象とした。これらの日本語機能表現について、文脈に依存せずモンゴル語訳が一意に決められる割合を調査した結果、級に応じて、73~94%という高い割合で一意に訳せることが分かった。さらに、この訳語対応を日本語からモンゴル語への機能表現翻訳規則とみなして、[グループ・ジャマシイ 1998]に

¹ モンゴル語で使われる文字はキリル文字である。モンゴル語では、自立語の語幹に対して格を表す語尾や動詞の活用を表す語尾・接続助詞等が結合したものが単語を構成し、ヨーロッパ言語と同様に、空白で区切られた単語の列により文を構成する。通常、名詞の語幹には、数を表す語尾、格を表す語尾、再帰・所有を表す語尾がこの順に接続する。動詞の語幹に接続する語尾は、命令・願望類、叙述類、完了・習慣等を表す類、順序関係を表す類、等に分類される[亀井 1992]。

含まれる 620 例文における機能表現の翻訳を行ったところ、完全正解率が 80%、部分正解率が 95%という高い精度が得られた。本論文では、これらの調査結果の詳細について述べ、日本語・モンゴル語間の機能表現の機械翻訳を実現するために、今後解決すべき課題について検討する。

2. 日本語・モンゴル語間の機能表現の対応付け

本節では、日本語の機能表現をモンゴル語に翻訳する方向について、日本語・モンゴル語間での機能表現の対応付けを行い、訳語対応の曖昧性の度合い、および、機能表現の内部構造の対応のバリエーション等を調査した結果を述べる。本論文では、日本語非母語話者にとっての難易度と翻訳の難しさの間の関係を考慮して、日本語能力試験における級別に機能表現を分類した。そして、日本語能力試験の 2~4 級の各級に対して、日本語能力試験出題基準[日本国際教育協会 2002]の文法項目において掲載されている機能表現 100~200 個(合計で 455 個)を対象として調査を行った。

2.1 モンゴル語訳の曖昧性の有無

まず、日本語の機能表現をモンゴル語に翻訳する方向において、全機能表現を、「モンゴル語側に対応する表現がない」、「日本語の機能表現の複数の意味・用法によって異なるモンゴル語訳が対応し、訳語選択が必要」、「モンゴル語訳の表現が一意に決定できる」の三種類に分類した。この内訳を表 1 に示す。機能表現の難易度が高いほど、一意にモンゴル語訳の表現が決まる割合が高くなり、2 級で 94%、3 級で 86%、4 級で 73%となった。2 級の機能表現は日本語の中級コースで習得するレベルの表現で、慣用的で用法が一つしかないものも多い。したがって、その多くは、文脈によらずモンゴル語訳が一意に決まった。一方、4 級の機能表現は、日本語の初級コースの前半で習得するレベルの基本的な表現で、複数の意味・用法を持つものが多く含まれている。このため、モンゴル語訳においても複数の表現に訳し分ける割合が多くなった。しかし、全体としては、日本語からモンゴル語の方向への翻訳においては、かな

表 1. 日本語機能表現: モンゴル語への
対応可能性の分類

級	モンゴル語 対応表現 なし	モンゴル語 訳 曖昧性あり	モンゴル語 訳 曖昧性なし	合計
2	0 0.0%	12 5.9%	193 94.1%	205 100.0%
3	2 1.7%	14 12.0%	101 86.3%	117 100.0%
4	8 6.0%	28 21.1%	97 72.9%	133 100.0%

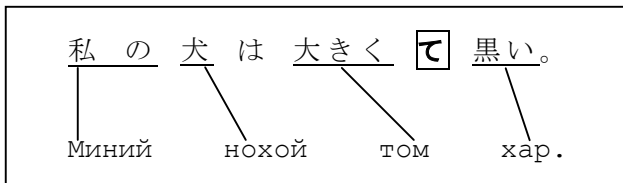


図 1. 表 1 「モンゴル語対応表現なし」の例
(機能表現 = 「て」)

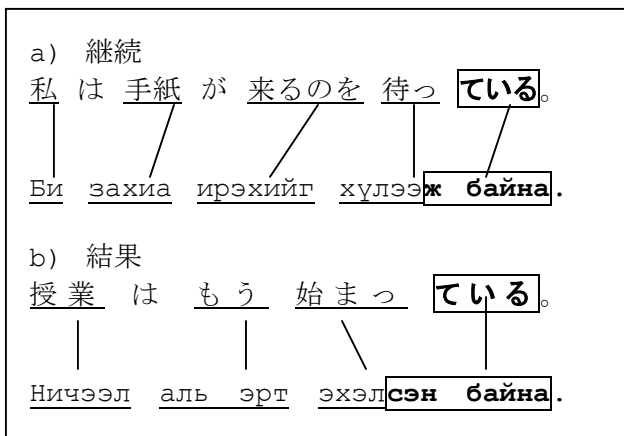


図 2. 表 1 「モンゴル語訳曖昧性あり」の例
(機能表現 = 「ている」)

り高い割合で、一意に機能表現の訳が決められる
と言える。

図 1 に、「モンゴル語対応表現なし」の例として、機能表現「て」の日本語例文中の各語をモンゴル語訳文の各語に対応付けたものを示す。モンゴル語の形容詞は基本的には活用しないため、日本語文中の接続助詞「て」は、モンゴル語には翻訳されない。形容詞「大きい」の活用形「大きく」(あるいは語幹「大き」)だけが「大きい」の意味のモンゴル語形容詞“том”に対応する。また、図 2 に、「モンゴル語訳曖昧性あり」の例として、機能表現「ている」の二つの意味(意味分類は、[グループ・ジャマシイ 1998]のものを採用)、および、それぞれの意味での「ている」の日本語例文とモンゴル語訳を示す。

次に、表 1 で「モンゴル語訳曖昧性なし」と分類された機能表現に対して、モンゴル語訳表現との間で一対一の関係が成り立つ(直訳・J=M)か、

表 2. 表 1 中「モンゴル語訳曖昧性なし」の分類

級	直訳		言い換え		合計
	J=M	J<M	J=M	J<M	
2	160 62.0%	33 12.8%	54 20.9%	11 4.3%	258 100.0%
3	85 67.5%	33 26.2%	6 4.8%	2 1.6%	126 100.0%
4	73 68.9%	24 22.6%	8 7.5%	1 0.9%	106 100.0%

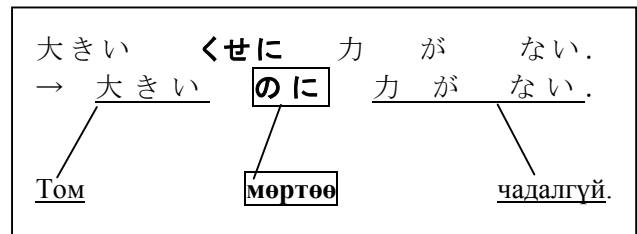


図 3. 表 2 「言い換え・J=M」の例
(機能表現 = 「くせに」)

それとも、モンゴル語訳表現の方が広い意味に対応していて、日本語表現はその一部の意味だけに対応する(直訳・J<M)のかの内訳を調査した。さらにここでは、そのままではモンゴル語に直訳するのが困難な日本語機能表現は、いったん別の機能表現に言い換え²てからモンゴル語機能表現に翻訳するという手順³をとった。言い換えられた先の機能表現がモンゴル語訳表現との間で一対一の関係を持つ(言い換え・J=M)か、それとも、モンゴル語訳表現の方が広い意味に対応する(言い換え・J<M)のかの区別をして、その内訳を調査した。この結果を表 2 に示す。これから分かるように、2~4 級では 70%近い機能表現が「直訳・J=M」に分類されており、モンゴル語から日本語の方向への翻訳においても、訳し分けが必要となる場合がそれほど多くないことが分かる。また、直訳が困難で言い換えが必要な割合は、2 級では約 25%と最も多いが、その多くは「言い換え・J=M」に分類されている。図 3 に、「言い換え・J=M」の例として、モンゴル語への直訳が困難な機能表現「くせに」(2 級)を「のに」(3 級)に言い換えた後、文全体をモンゴル語に翻訳したものを示す。

² モンゴル語への直訳が困難な日本語機能表現については、原則として、同一級内もしくはより平易なレベルの級の機能表現で意味の近い表現に言い換えてからモンゴル語に訳すこととした。ただし、日本語能力試験出題基準[日本国際教育協会 2002]に掲載されていない表現への言い換えが必要な場合もあり、その場合は、言い換え先の級に関する制約は無視した。

³ この方法は、機械翻訳において、原言語の表現の全体を、目的言語への翻訳が容易な基本表現集合とその他の表現の集合に分け、基本表現以外の表現はいったん基本表現に言い換えてから、基本表現の翻訳規則を流用するというアプローチ[Yamamoto 2002]に基づいている。

表 3. 表 1 中「モンゴル語訳曖昧性なし」における多義性保持に関する分類

級	多義性保持	多義性なし	合計
	45	136	181
2	24.9%	75.1%	100.0%
3	28	60	88
	31.8%	68.2%	100.0%
4	12	63	75
	16.0%	84.0%	100.0%

表 4. 表 3 中「直訳・J=M」における構成的対応可能性の分類

級	構成的	一部構成的	1 単語対 1 単語	非構成的	合計
2	34	32	4	90	160
	21.3%	20.0%	2.5%	56.3%	100.0%
3	20	14	18	33	85
	23.5%	16.5%	21.2%	38.8%	100.0%
4	9	13	31	20	73
	12.3%	17.8%	42.5%	27.4%	100.0%

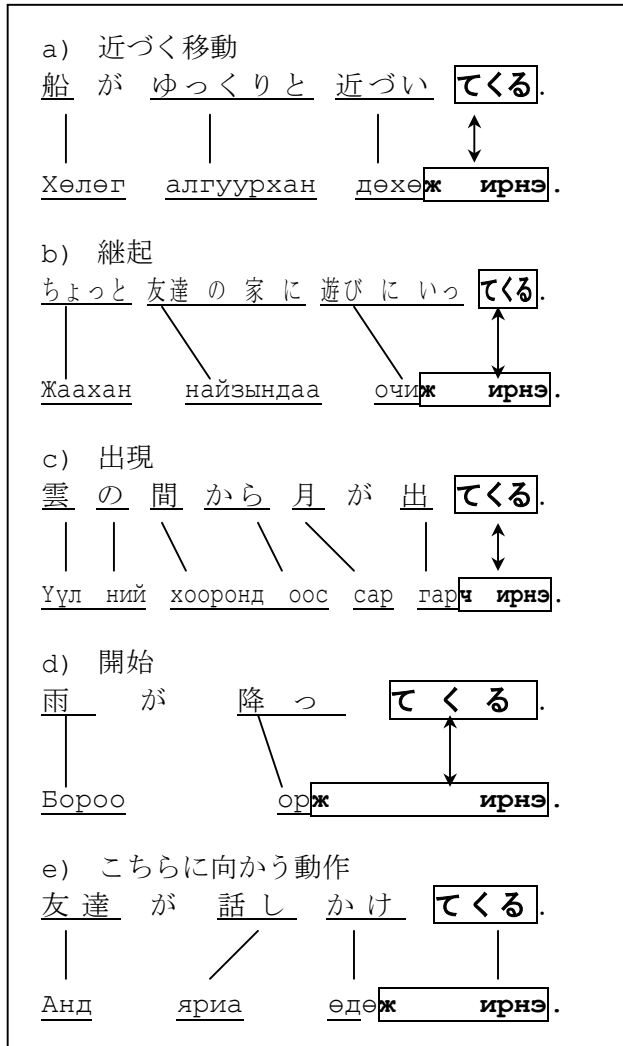


図 4. 日本語機能表現「てくる」のモンゴル語訳における多義性保持

さらに、表 1 で「モンゴル語訳曖昧性なし」と分類された機能表現に対して、実際には、日本語機能表現には複数の意味・用法があるにもかかわらず、その多義性を保持したまま一つのモンゴル語機能表現に翻訳することが可能な場合(多義性保持)、および、そもそも日本語機能表現の方には

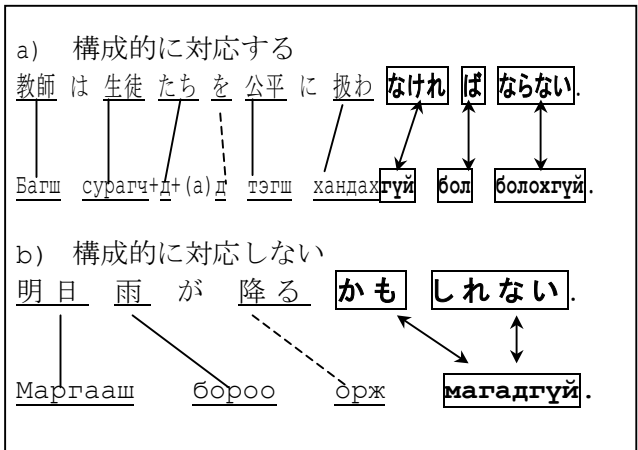


図 5. 表 3 中「直訳・J=M」における構成的対応可能性の例

一つの意味・用法しかない場合(多義性なし)の分類を行った。ただし、ここでは、一つの機能表現の持つ全ての意味・用法をただ一つのモンゴル語機能表現に翻訳できる場合のみを、「多義性保持」と分類した。この内訳を表 3 に示す。この結果から、各級に応じて、16~32%の機能表現について、その多義性を保持したままモンゴル語機能表現に翻訳できることが分かる。図 4 に、機能表現「てくる」の五つの意味・用法(意味分類は、[グループ・ジャマシイ 1998]のもの)の例文、および、それらの例文のモンゴル語訳を示す。

2.2 機能表現の構成的対応

表 2 で「直訳・J=M」(モンゴル語訳表現に直訳でき、一対一の関係が成り立つ)と分類された機能表現に対して、対応するモンゴル語機能表現の内部構造との間で、内部構造が構成的に対応可能であるかどうかを調査した。ここでは、構成的対応可能性を大きく以下の四通りに分類した。i) 日本語・モンゴル語のどちらの機能表現も $n(n \geq 2)$ 単語から構成され、しかも、順序を保持して構成単語が一対一に対応する(構成的)、ii) 一部の構成単語が対応する(一部構成的)、iii) 日本語・モンゴル語のどちらの機能表現も一単語から構成される(一単語対一単語)、iv) その他(非構成的)。この内

訳を表 4 に示す。全体では、各級に応じて、30～40%の機能表現が(一部)構成的に対応していることが分かる。図 5 に「構成的」な対応の例、および、「非構成的」な対応の例を示す。構成的に対応する機能表現が一定の割合で存在することからも分かるように、日本語とモンゴル語は文法および語の構造がよく似ていると言える。また、4 級では 43%、3 級では 21%の機能表現が一単語対一単語対応となっているのに対して、2 級で一単語対一単語対応となる機能表現はごくわずかである。これは、2 級の機能表現においては、「なければならない」のように、機能表現全体の意味が構成的に合成できない慣用的な表現が多くを占めるためである。

表 5. 機能表現の翻訳性能の評価
(30 機能表現, 機能表現単位)

級	正解	部分正解	誤り
2	89.9%	8.6%	1.5%
3	75.7%	18.9%	5.4%
4	82.4%	11.3%	6.2%
全体	82.7%	12.9%	4.4%

表 6. 機能表現の翻訳性能の評価(例文単位)

級	正解/全体	部分正解/全体	誤り/全体
2	146/163 (89.6%)	15/163 (9.2%)	2/163 (1.2%)
3	186/251 (74.1%)	50/251 (19.9%)	15/251 (6%)
4	162/206 (78.6%)	29/206 (14.1%)	15/206 (7.3%)
全体	494/620 (79.7%)	94/620 (15.2%)	32/620 (5.2%)

3. 日本語からモンゴル語への機能表現の翻訳

表 1 で「モンゴル語訳曖昧性なし」と分類された機能表現とそのモンゴル語訳表現との間の訳語対応を、日本語からモンゴル語への機能表現翻訳規則とみなして、その翻訳性能を評価した。ここでは、2～4 の各級に対して、機能表現を 30 個選び⁴、これらの機能表現について、[グループ・ジャマシイ 1998]に含まれる全例文(合計で 620 文)における機能表現の翻訳を行い、その精度を評価した。精度の評価は「正解」「部分正解」「誤り」の三段階で行い、機能表現ごとに算出した精度を全機能表現について平均した結果を表 5 に、また、例文単位での平均により算出した精度を表 6 に、それぞれ示す。いずれの評価においても、完全正解率が 80%以上、部分正解率が 95%という高い精度を達成した。また、表 2 において、直訳・言い換えを総合した「J=M」(機能表現の対応が一對

一)の割合が相対的に高い 2 級の完全正解率が 90%となり、3・4 級よりも高い精度となった。以上の結果から、日本語の機能表現のうちの多くを占める「モンゴル語訳曖昧性なし」の機能表現について、モンゴル語への機械翻訳が実際にかなり高い精度で行える可能性があることが分かった

4. おわりに

本論文では、日本語の機能表現をモンゴル語に翻訳する方向について、日本語・モンゴル語間での機能表現の訳語対応の曖昧性の度合いを調査した。その結果、日本語機能表現の多くについて、文脈に依存せずモンゴル語訳が一意に決められることが分かった。さらに、それらの日本語機能表現をモンゴル語へ翻訳する規則の評価を行い、完全正解率が 80%、部分正解率が 95%という高い精度を達成した。今後は、モンゴル語への翻訳の際に訳し分けが必要となる日本語機能表現について、訳語選択規則を作成し、日本語機能表現全体についてモンゴル語への機械翻訳を実現する。訳し分けが必要な機能表現の中には、格助詞等の頻出する表現も含まれているので、これらの訳語選択規則の実現が重要となる。

今後、日本語・モンゴル語間の機械翻訳全体を実現するためには、その他に、まず、自立語に関する日本語・モンゴル語対訳辞書が必要となる。これについては、すでに、見出し語数 6,000 語程度の規模のオンライン辞書を入手済みであり、当面はこれを利用する。さらに、自立語に関する訳語選択の実現、モンゴル語から日本語への翻訳における機能表現翻訳規則の作成、モンゴル語の語構成規則の作成、および、それをを用いたモンゴル語句生成・形態素解析の実現、といった要素技術を確立する必要がある。今後は、これらの課題に取り組む予定である。

参考文献

- [亀井 1992] モンゴル語, 亀井孝, 河野六郎, 千野栄一(編著), 言語学大辞典, 第 4 巻, 世界言語編(下-2), 三省堂, pp. 501-517, 1992.
- [グループ・ジャマシイ 1998] グループ・ジャマシイ: 教師と学習者のための日本語文型辞典, くろしお出版, 1998
- [日本国際教育協会 2002] 日本国際教育協会: 日本語能力試験出題基準(改訂版), 国際交流基金(財) 日本国際教育協会, 凡人社, 2002.
- [Yamamoto 2002] K. Yamamoto: Machine Translation by Interaction between Paraphrase and Transfer, *Proc. 19th COLING*, pp. 1107-1113, 2002.

⁴ これらの機能表現は、表 2 の 12 分類における各分類から、表 2 の機能表現数の分布をほぼ忠実に反映するように選んだ。