

日韓機械翻訳における否定文の処理

2E-6

金 泰錫 浦 昭二

慶應義塾大学理工学研究科

1. はじめに

否定を表す「ない」、「ん(ぬ)」が使われている日本語の述部は、韓国語では、否定表現(否定素と否定補助用言)を用いた否定文、または対立語を用いた肯定文に訳される。しかも、丁寧と否定の意味を表す助動詞が使われた日本語の述部とは、表現順序が異なる。これらの違いについては、文献2)で考察を行った。本論文では、翻訳テーブルを用いた日韓機械翻訳システムに^{1,2)}、以下の処理を補完し、否定語が使われている日本語の述部の翻訳処理を試みた。(1)翻訳する文の形式を判別する。①否定文の場合は、適切な韓国語の否定補助用言を選択するルールを用いる。②肯定文の場合は、対立語を用いた肯定文に変換するルールを用いる。(2)両国語間の表現順序の違いが生じる場合は、前接単語の翻訳テーブルを利用し、その表現順序を合わせる。

その結果、日本語の否定文を自然な韓国語に翻訳することが可能になり、日韓機械翻訳にとって大きな問題であった述部の翻訳精度をより向上できた。

2. 日本語の否定文の翻訳方法

2.1 訳文種類の決定

日本語の否定文は韓国語の否定文、または肯定文に訳される。前者は、韓国語の否定補助用言を用いる場合であり、後者は、否定された動詞の否定の意味を表す対立語を用いる場合である。しかし、対立語を用いる表現は、心の状態や存在の有無についての韓国語の独特な表現の仕方であるため、これらの動詞についてはあらかじめ辞書に、その存在を登録し、その有無によって訳文の種類を区別する。

2.2 否定補助用言の選択方法

日本語の否定語に対応する韓国語の否定補助用言「지 않다(ji anta)」と「없다(opda)」は、文が表す否定種別によって使い分けられる。否定文は、その文が表す意味から単純否定、意図否定、能力否定の3種類に分けられる。これらを否定種別と呼ぶ。日本語の否定文を翻訳するためには、文が表す否定種別を判別し、それに応じた韓国語の否定補助用言を用いる必要がある。否定種別は、否定語の前にくる助詞の区分、または用言などの判別によって判断可能である。否定補助用言を選択するルールを表1で示す。

2.3 否定語の翻訳テーブル

韓国語は、一般的な韓国語の音韻規則では説明できない活用を行なう用言(不規則用言)が多く、これらに対して一般的なルールを作成することは大変難しい。韓国語の否定補助用言についても、日本語の述部の接続情報(否定文に後接する別の助動詞の意味や品詞)を考慮した適切な訳語(活用語尾、または必要によって連結語尾などを付けた訳語)をあらかじめテーブル形式でまとめたものを用いることは、翻訳処理において効率的である¹⁾。否定補助用言「없다(opda)」の翻訳テーブルを表2に示す。「지 않다(ji anta)」の翻訳テーブルも表2と同様に作成した。

表1 韓国語の否定補助用言の選択ルール

選択ルール1: n1 ¹⁾ +助詞(が、は、も、の、こそ、さえ、しか、までだけ)+否定語(「ない」のみ)	⇒「없다(opda)」(単純否定)
選択ルール2: n2 ¹⁾ +助詞(では、じゃ、でも)+否定語(「ない」のみ)	⇒「지 않다(ji anta)」(単純否定)
選択ルール3: 可能を表す動詞 ²⁾ +否定語	⇒「없다(opda)」(能力否定)
選択ルール4: 可能を表す動詞以外の動詞+否定語	⇒「지 않다(ji anta)」(単純否定、または意図否定)
選択ルール5: 可能の意味(れる、られる)+否定語	⇒「없다(opda)」(能力否定)
選択ルール6: 形容詞+否定語	⇒「지 않다(ji anta)」(単純否定)
選択ルール7: 形容動詞+否定語	⇒「지 않다(ji anta)」(単純否定)

注1) n1は、体言、または形容動詞を表す。n2は体言、形容詞、形容動詞のいずれかを表す。

注2) 可能を表す動詞とは、5段動詞から転じた下1段動詞
[例:書く⇒書ける]と可能の意味を表す動詞「出来る」を示す。

表2 日本語の否定語の翻訳テーブル(否定補助用言「없다(opda)」の場合)

活用形	活用単語	後接意味及び文法情報による韓国語の対応		例文	
		No	否定語の後接情報		韓国語の対応
未然形	なから	1	推定、意志を表す助動詞	없(op)	漢字が書けなからう。
連用形	なかつ	2	過去を表す助動詞	없었(opseut)	漢字が書けなかった。
	なく	3	連用形に接続可能な助動詞	없어(opeu)	漢字が書けなくても。
	ず	4	連用中止形(,)	없고(opego)	漢字が書けず。
終止形	ない	5	終了	없다(opda)	漢字が書けない。
	ん(ぬ)	6	丁寧を表す助動詞	없(op)	漢字が書けないでしょう。
連体形	ないん(ぬ)	7	終止形に接続可能な助動詞、または疑問、推定を表す助動詞	없(op)	漢字は書けないが、漢字が書けないそうだと、漢字が書けないらしい。
		8	連体形に接続可能な助動詞、または体言	없는(opeun)	漢字が書けないのに漢字が書けない人は
仮定形	なけれぬ	9	仮定を表す助動詞	없으(opeu)	漢字が書けなければ、漢字が書けねば。

2.4 対立語を用いた肯定文

韓国語では、心の状態、または存在の有無を表す動詞の否定は、その否定の意味を持つ対立語で表す。この場合、日本語の否定文が韓国語では、肯定文に訳される。そのため、文の形式を肯定文に変換するルールが必要である。その交換ルールを表3に示す。

2.5 表現順序の違い

述部に否定を表す「ん」と丁寧を表す助動詞「ます」が同時に使われている日本語の述部「ません」は、韓国語の述部とその表現順序が異なる²⁾。本論文で試みている日韓機械翻訳システムでは、各助動詞ごとに翻訳テーブルを用意している。つまり、助動詞「ます」の翻訳テーブルに、未然形の「ませ」の訳語として、「否定+丁寧」の意味を表す訳語をあらかじめ登録することによって、表現順序の違いに対処することにする。

表3 対立語による述部の変換ルール

変換ルール1: 用言 + 否定語 ⇒ 対立語 (未然形) (n1形) (n1形)
変換ルール2: 用言 + 丁寧 + 否定語 ⇒ 対立語 + 丁寧 (連用形) (未然形) (n2形) (連用形) (n2形)
変換ルール3: 用言 + 丁寧1 + 否定語 + 丁寧2 (連用形) (未然形) (終止形) (n3形) ⇒ 対立語 + 丁寧1 (連用形) (n3形)

注) n1形、n2形、n3形は、各活用形を表わす。

3. 否定文の翻訳手順

3.1 否定補助用言を用いた翻訳

日本語の否定語が韓国語の否定補助用言を用いて翻訳される場合は、「지 않다(ji anta)」、または「없다(opda)」のどちらが適切かを選択しなければならない。

例文1) 1つしかない地球環境を守らない。
1개밖에 없는 지구환경을 보호하지 않다.

訳文1) (Igebakeopnun jiguhwanyongul bohohaji anta.)
例文1) で、「1つしかない」の「ない」の場合は、その前接助詞が「副助詞(しか)」であることから、表1の否定補助用言の選択ルール1番によって、単純否定を表す「없다(opda)」が選択される。次に、「ない」の後接単語が「地球(体言)」であることから、「ない」の活用形は連体形である。したがって、表2の翻訳テーブルの8行目の「없는(opnun)」に訳される。そして、「守らない」の場合も同様に処理する。

3.2 対立語を用いた翻訳

例文2) 見るかも知れません。

訳文2) 볼지도 모릅니다. (boljido morumnida.)
日本語の動詞「知れる」に対応する韓国語は「알다(alda)」である。「知れる」の連用形「知れ」の翻訳テーブルには、否定語が後接した場合に、その対立語が存在することを示すマーカー(辞書には、対立語マーカー“*”とその対立語)が表示されている。そして、否定の意味を持つ助動詞「ん」は対立語の「모르다(moruda)」の中にその意味が含まれるために、肯定文に変換する必要がある。例文2)の場合は、述部形態と一致する表3の2番(n2=終止形)が適用され、肯定文に変換される。

4. 翻訳実験と考察

実験には、文献4)の論文(約14万文字)の中から、否定を表す「ない」、「ん(ぬ)」が使われている否定文をすべて取り出し、翻訳処理を行った。表4で示すように、全体で否定文は548個あったが、その約62%が否定補助用言を用いた否定文に訳された。そして、表5に、否定語の翻訳テーブルの訳語別の頻度を示す。否定補助用言を用いた否定文に訳された342個が、2つの否定語の翻訳テーブルに用意した訳語によって翻訳可能であったことから、これらの翻訳テーブルが有効であると思われる。また、その中で、表現順序の違いが生じた例文は、「加えません。」などの17文のみであり、その数は少なかったが、十分対応できた。

そして、対立語を用いた肯定文への翻訳は、約34%(185個)を占めている。表6に、185個の文がその述部を肯定文に変換するために適用された表3の変換ルールの適用頻度を示す。対立語を用いる肯定文に翻訳するために用意した表3の変換ルールで十分であった。

そして、処理できなかった例文としては、ことわざの「人間はパンのみに生きるものにあらず。」であり、科学技

術文献を対象とする本論文では、きわめて特殊な表現と考えられる。否定語が含まれている日本語の述部548文を対象として、翻訳処理を行った結果、人間が読んで理解できる範囲であり、なおかつ自然な表現になっているものを合格として、約99.8%の合格率を得ることができ、満足な翻訳結果が得られたものといえる。

表4 翻訳形態別の頻度

No	翻訳形態の区分	頻度
1	否定補助用言を用いた肯定文	238 (43.4%)
	"지 않다(ji anta)"	104 (19.0%)
2	対立語による肯定文	185 (33.8%)
3	慣用的な表現	20 (3.6%)
4	処理出来なかった文(ことわざ)	1 (0.2%)
合計		548 (100%)

表5 韓国語の否定語補助用言が用いられたときの訳語別頻度

否定語の活用形	No	-지 않다(ji anta) が用いられた頻度	- 없다(opda) が用いられた頻度
未然形	なから	1 (0.4%)	1 (1.0%)
	なかつ	2 (0.8%)	5 (4.8%)
連用形	なく	3 (2.1%)	2 (1.9%)
	ず	4 (21.0%)	17 (16.3%)
終止形	ない	5 (6.3%)	14 (13.5%)
	ん(ぬ)	6 (6.3%)	9 (8.6%)
		7 (21.9%)	19 (18.3%)
連体形	ないん(ぬ)	4 (17.2%)	3 (30.8%)
仮定形	なければ	5 (24.0%)	5 (4.8%)
合計		238 (100%)	104 (100%)

表6 対立語による述部の変換ルールの適用頻度

述部変換ルール	用いられた頻度	
変換ルール1	未然形	0 (0%)
	連用形	18 (9.7%)
	終止形	41 (22.2%)
	連体形	22 (11.9%)
	仮定形	3 (1.6%)
変換ルール2	終止形	89 (48.1%)
	連体形	0 (0%)
変換ルール3	未然形	0 (0%)
	連用形	0 (0%)
	終止形	2 (6.5%)
	連体形	0 (0%)
合計	185 (100%)	

5. 終わりに

日本語の否定語が使われている述部を韓国語に翻訳するため、著者らが提案した翻訳テーブルを用いた日韓機械翻訳システムに、さらに上記の処理を補完した結果、546個の否定文がほぼ完全な韓国語に翻訳することが可能であった。しかし、日韓機械翻訳システムを実用化させるためには、単語の意味判別や用言の多意性などの問題が残されている。

参考文献

- 1) 金泰錫, 浦昭二: 意味接続関係に基づいた翻訳テーブルを用いた日韓機械翻訳システム、情報処理学会第43回全国大会(3), pp. 203-204(1991).
- 2) 金泰錫, 浦昭二: 日韓機械翻訳における述部の否定語の処理、情報処理学会第44回全国大会(3), pp. 141-142(1992).
- 3) P. W. Anderson (訳: 森、米沢): 科学の美学について、ヒューマン・ハーモニー、三田出版会(1989).