

日中機械翻訳における存在文の翻訳処理について

王軼謳 ト朝暉 宇野修一 浅井良信 池田尚志
岐阜大学工学部

1 はじめに

存在文はどの言語でも基本的な表現の種類の一つであり、それぞれの言語で特徴があり言語により異なりが現れてくる。日本語と中国語でも、前者が存在の主体が有情物か非情物かで使われる動詞が異なる(「ある/いる」)のに対し、後者では所在の意味か所有の意味かで使われる動詞が異なる(「在/有」)など、大きな違いがある。日本と中国の言語学の分野では存在文について論述があるが[1][2]、日中機械翻訳の角度からの研究は殆ど見あたらない。また現在の日中市販翻訳ソフトでは、存在文に関する誤りが多く見られる。我々は日本語文の構文特徴、中国語文の構文特徴、対応名詞の意味属性などを利用して存在動詞の翻訳規則をまとめ、存在文の翻訳方法について提案した[4]。本論文では翻訳規則をさらに整備して我々が開発している翻訳システムで、及び手作業で翻訳実験を行い、問題点を分析した。

2 「ある」と「いる」について

本節では「ある」、「いる」、「ない」を対象として、日中両言語における存在動詞の翻訳規則を中心に述べる。

2.1 「ある」(非情物存在主文)について

「ある」は中国語に翻訳すると、意味上の「所在」か「所有」に焦点を置くことによって、基本的に「有」か「在」で表現し、ほかに場合によっては「在于、发生、进行」などの別の動詞で対応することもある。

「ある」の意味分類を行って、日本語文の構文特徴(文型、助詞、テンスなど)と存在主と存在場所の意味属性、構文的な特徴を用いて、中国語の存在動詞の組み合わせの制限を総合して分析し、判定条件を機械で処理可能な形に整理した[3](表1にその一部を示す)。表中のN1、N2、N3は名詞を、Xは単語、句、文などを、Pは結び部分(述部)を表す。

表1 「ある」の翻訳規則(一部分)

意味分類	判定の条件とその属性					中訳語
	日本語パターン	名詞N1の属性	名詞N2の属性	名詞N3の属性	Xの構造と成分	
1	(1)	N1はN2にある	具体/固有名詞	場所/場/地名/組織名		N1位于/在N2
2	(2)	N1はN2にある	抽象	具体		N1在N2
3	(3)	N1はN2にある	抽象	関連/性質/状態/存在		N1在于N2
4	(4)	N1に/はN2がある	場/具体/固有名詞	具体/固有名詞		N1有N2
5	(5)	N1に/にはN2が/の/NULLある。	体言	抽象		N1有N2
	(6)	N1はN2に/に対してN3がある	主体	体言	抽象	N1对N2有N3
6	(7)	N1でN2がある	場所/場	式、行事		在N1举行N2
	(8)	N1でN2がある	場所/場	相談		在N1进行N2
7	(9)	N1に/はXことがある	主体		動詞連体形過去形	N1曾经X
8	(10)	N1でN2がある	場所/場	来事/天災		在N1发生N2

表1のような「ある」の翻訳パターンは26個ある。そのほかに、個々の語彙そのものを制約条件とする（字面照合）のが適切なパターンがある（現在のところ33個）。例えば

（N1：人/抽象物/性質）には表裏がある

→ N1 表里不一

これらを含めると「ある」の翻訳パターンとして全体で59個を設定した[3]。

2.2 「いる」（有情物存在主文）について

「いる」は中国語に翻訳すると、基本的に「有」か「在」で表現し、ほかに場合によっては「呆、处于…」などの別の動詞で対応することもある。「ある」の場合と同様に、「いる」の意味分類を行って、翻訳規則をまとめて、現在のところ16個を設定した[3]。

2.3 存在否定文の翻訳処理

述語が「ある」の否定「ない」と、「いる」の否定「いない」等で構成される文を存在否定文という。

「いる」の否定「いない」の存在否定文の翻訳処理は通常の動詞の否定処理[4]（「いる」の翻訳プラス否定の翻訳）でできる。しかし、「ない」では単純に「ある」の否定としては翻訳できない場合もあるので、我々は「ない」を一つの自立語として、翻訳規則を整理した（今のところ45個）[3]。

翻訳規則と照合できない「ない」は、「ある」と対応できると考えて、存在文「ある」の翻訳プラス否定の翻訳という方法で翻訳する。

3 翻訳実験と評価

3.1 Jaw/Chinese での翻訳実験

評価に用いたデータは日英対訳コーパス[5]の約15万文の初めの2万文中から抽出した存在文である。この中に、「ある」、「いる」、「ない」を持つ文はそれぞれ989、379、832文が存在した。更に重複している文と存在表現ではない文を除外し、残りの917、305、787文を存在文の例文とした。このうち、「ある」の400文、「いる」の100文と「ない」の100文を翻訳規則を作るための調査対象として観察し、分析した。

2節で述べた翻訳規則を我々の研究室で開発している日中機械翻訳システム Jaw/Chinese[6][7]に組み込んで、調査対象文中の「ある」の72文と「いる」の18文をJaw/Chineseで文の全体を翻訳した。クローズテストであるが、良好な翻訳結果を得た。

3.2 手作業による翻訳実験と評価

存在文の例文の中で調査対象としなかった文の中から、「ある」273文、「いる」177文、「ない」191文を抽出し、2節の翻訳規則を用いて手作業で翻訳実験を行った（オープンテスト）。個人判断（存在表現の翻訳に着目し、その訳語と語順と助詞の翻訳が合っているかどうかという観点から）で評価を行った。同時にある市販機械翻訳ソフトを用いて翻訳し、その結果も評価した（表2）。

表2 例文翻訳結果（手作業：オープンテスト）

	2万文中の存在文	評価文	正訳文		正訳率	
			A	B	A	B
「ある」	917	273	261	167	95%	61%
「いる」	305	177	165	57	93%	31%
「ない」	787	191	172	80	90%	42%

（A：我々の訳 B：市販ソフトの訳）

評価した際、我々の規則の評価については存在表現の部分以外の翻訳は全部正しく翻訳されたものとして、存在表現の翻訳に関する我々の規則を適用した結果のみを評価しており、一方市販ソフトに対する評価では、存在表現の部分以外の翻訳の正否は無視して、存在部分の正誤のみに着目して評価している。以下に手作業での翻訳例を示す。

「この金貨は古銭を集めている人にととても価値がある。」

step1 構文解析し、存在動詞に関わる下記の各情報を得る。

存在動詞＝「ある」。

「金貨」の意味属性は「貨幣」で、「は」が付いている。

「人」の意味属性は「人間」で、「に」が付いている。

「価値」の意味属性は「是非」で、「が」が付いている。

「とても」は程度副詞であり、「ある」を修飾している。

日本語文の構造は「N1はN2にN3がある」である。

Step2 Step1の結果と「ある」の翻訳規則を照合する。

表1の(6)と照合できる。

中国語文の構造は「主語＝货币、謂語＝有、目的語＝价值、対象者＝人、対象者の修飾語＝对」である。

Step3 「とても」の翻訳結果と基本部分の訳とをあわせて生成し、通常の文の線状化の語順によって、下記の訳文を得る。

→中国語： 银币对人很有价值。

評価の結果から見ると、全体として90%以上の正訳率であり、市販の翻訳ソフトの現状と比較すると、我々の正訳率は各々34%、62%、48%まさっており、我々の方法は十分な有効性が期待できると考えている。

同時に、翻訳規則に関して文型の影響（日本語入力文中での「ある」の位置）が問題になるかどうかについて、Jaw/Chineseで翻訳して検討した。

文型（1）「ある」は文末にある。

日本語：その教会は市を見下ろす丘の上にある。

翻訳文：那个教会在俯视城市的山冈上面。 (○)

文型（2）「ある」は文中の接続詞の前にある。

日本語：急ぎの用があるのてきょうはこれで失礼します。

翻訳文：因为有紧急的事情所以今天就此告辞 (○)

文型（3）「ある」は埋め込み連体修飾として使われている。

日本語：それは地域の人々にとって意義ある事業だ。

翻訳文：那是对地区的人们有意义的事业。 (○)

以上のように調査範囲内では、「ある」の位置によって、翻訳規則に変更を加える必要があるというような状況はなかった。

4 誤り分析

規則の評価の中で43文の誤りがあった。以下にそれらの誤訳原因等についての分析と考察を述べる。

A 規則の不足の問題

➤ (1)：突然言いようのない恐怖感に襲われた。

現訳：突然，没有说法的恐怖袭来。 (×)

正訳：突然，无法形容的恐怖袭来。 (○)

分析：「言いようのない」では、「无法形容」が正解だが「没有说法」に訳している。「言いようのない」

は慣用句であり、中国語でも「无法形容」は慣用句である。「ない」の規則に加えて「言いよう+の/が+ない」のようなパターンの追加が必要である。

➤ (2)：寒くて足の指の感じがない。

現訳：由于冷，没有脚趾的感觉。 (×)

正訳：由于冷，脚趾没有感觉。 (○)

分析：この文では、「ない」の訳語の選択（没有）は正しいが、中訳文の意味は間違っている。現訳の意味は「寒くて足の指がないと感じる。」である。「N（動物部分）の感じがある」というパターンを追加する必要がある。

➤ 「…つもりでいる」というパターンに対する誤訳。

(3)：親は彼を医者にするつもりでいた。

正訳：父母本来打算让他当医生的。

(4)：彼女はまるで小説のヒロインにでもなったつもりでいる。

正訳：她简直认为(自己)成了小说的女主角。

(5)：きみには来年主将になってもらうからそのつもりでいてくれ。

正訳：(我们)希望你明年成为主将，为此(你)要做好准备啊！

(6)：彼は自分がチームのエースのつもりでいる。

正訳：他认为自己是球队的主攻投手。

分析：元の規則には「…つもりでいる」という構造に対するパターンが無かったために翻訳できなかった。「…つもりでいる」には対応する中国語の意味がいくつかあるので分析が必要である。今のところ

①「N1(人)は +Vる (動詞連体形現在式) +つもりでいる」、

②「N1(人)は +Vた (動詞連体形過去式) +つもりでいる」

③「Nのつもりでいる」

という規則を増やすことで文3、文4と文6に対応できると考えている。文5では「…そのつもりでいる」という字面照合パターンで翻訳するのが適切であろう。

このように「ある」、「いる」と「ない」に関しては慣用的/固定的な表現が多い(誤訳43文のうち32文はそのような表現であった)。さらに多くの文例を翻

訳しながら、そのようなデータを集積、整理していくことが必要である。

B 「いる」の翻訳と副詞の問題

(7) : 私は彼に1日いてくださいと頼んだ。

現訳: 我请求他在一天。(×)

正訳: 我请求他呆一天。(○)

(8) : 私は彼にもう1日いてくださいと頼んだ。

現訳: 我请求他再在一天。(×)

正訳: 我请求他再呆一天。(○)

(9) : 彼は10年前はボストンにいた。

現訳: 他10年前在波士顿。(○)

誤訳: 他10年前呆波士顿。(×)

分析: 文7の「1日」は「数量詞+時間名詞」の形をとり、時間をあらわし、中国語では、補語になって、動詞の後に置く。「いる」の訳語「在」は直接に補語とつながるのは不自然であり、「呆」に翻訳される場合は自然である。「在」と「有」の後に各種の補語を用いることができない。また、文8の「もう」の「再」と「いる」の「在」は中国語の同音語であって、音の異なる「呆」を選択する方が自然な翻訳である。文7と逆に、文9の「10年前」は、中国語では、状語として、動詞の前に置く。この場合には「いる」は「在」に表現するのが正しく、「呆」に翻訳するのは正しくない。このように「いる」の中訳は時間副詞と関連して訳し分けが必要である。時間表現の分類、時間副詞と動詞の位置関係などさらなる考察が必要である。

C 存在主省略の翻訳問題

(10) : ここにイスが6つあり隣の部屋にはもっとある。

現訳: 这里有6把椅子, 在隔壁的屋子更多。(△)

正訳: 这里有6把椅子, 隔壁的屋子有更多。(○)

分析: 文10では前半の文で、存在主を主語として言及しているが、後半の主文では省略している。現在の規則では前半の文「ここにイスが6つある」の「ある」は「有」に翻訳されるが、後半の文「隣の部屋にはもっとある」の「ある」は「在」に翻訳される。後半の文だけが独立してあるのならこの訳でもよいが、この場合、主文の動詞としては前半の動詞と同じ「有」を用いるのが適切である。このよう

に、文脈上のことを考慮に入れた翻訳処理が必要になってくる。

D 機能語によって存在動詞の訳語が変わる問題

(11) : 一人でいる。

現訳: 一个人在。(○)

誤訳: 一个人呆。(×)

(12) : 一人でいたい気分だった。

現訳: 想一个人在。(×)

正訳: 想一个人呆着。(○)

分析: 用言後接機能語「たい」が付くと、「いる」の訳語は「在」から「呆」に変える必要がある。このように、機能語の影響も翻訳処理にとり入れる必要が出てくる。

5 おわりに

日中両言語の存在文における異同について考察し、日中機械翻訳のための存在動詞の翻訳規則を作成した。我々の研究室で開発している日中機械翻訳システム Jaw/Chinese に組み込んで翻訳実験し、また手作業による翻訳実験も行った。誤った翻訳例について誤訳の原因を分析した。

今後例文を更に増やして翻訳実験を進め、翻訳規則の整備と翻訳システムの改良を行っていく予定である。

参考文献

- [1] 飯田隆: 存在と言語—存在文の意味論, <http://phil.flet.keio.ac.jp/person/iida/papers/Sonzai.pdf>, 2001.
- [2] 劉月華, 潘文娛, 故鞏: 現代中国語文法総覧, くろしお出版, 1996.
- [3] 王軼謳, 卜朝暉, 宇野修一, 浅井良信, 池田尚志: 日中機械翻訳における存在文および関連する問題について, 情報処理学会研究報告 2006-NL-171, pp 95-102, 2006.
- [4] 卜朝暉, 池田尚志: 日中機械翻訳における否定文の翻訳, 自然言語処理 Vol.11, No.3, July.2004, p.97-112, 2004.
- [5] 村上仁一: 日英対訳データベースの状況. 「言語、認識、表現」第7回年次研究会, 2002.
- [6] 宇野修一, 福本真哉, 田中伸明, 松本忠博, 池田尚志: 日本語から多言語への翻訳エンジン jaw, 言語処理学会第11回年次大会発表論文集, 2005.
- [7] 謝軍, 今井啓允, 池田尚志: 日中機械翻訳システム jaw/Chinese における変換・生成の方法, 自然言語処理 Vol. 11, No. 1. p. 43-80, 2004.