

日中機械翻訳における名詞文について

陳 麗 山本いづみ

名古屋工業大学大学院工学研究科情報工学専攻

Chen Li Izumi Yamamoto

Department of Computer Science and Engineering

Graduate School of Engineering Nagoya Institute of Technology

1. 研究目的

日本語の「N1はN2です(N1, N2は名詞)」という名詞文は最も基本的な文型である。大部分の日本語初級教科書の冒頭に置かれる文型は「N1はN2です」である。しかし、現有の市販翻訳システムでは、名詞文の翻訳に誤訳が多く見られる。本研究では、日本語の「N1はN2です」という文型に注目し、機械翻訳の立場から、日中両言語の対照研究を行い、より精度の高い翻訳システムを実現するために有効な翻訳規則を提案することを目的としている。

日中機械翻訳における従来の手法では、日本語名詞文の対応文型は中国語の「是」字文であり、翻訳精度も低かった。そこで、研究者は、「是」字文以外で対応する例を分析し、その翻訳規則の概略について提案した(参考文献7)。本稿ではさらに用例数を増やして翻訳実験を行い、そこに現れた問題点について分析する。特に、「N1はN2です」の「N1」と「N2」の意味関係を考え、既提案の手法に必要な修正を加えて、より精度の高い翻訳規則を提案する。これは、日中機械翻訳における名詞文の翻訳精度向上に資するものであり、その意味は大きいと考える。

2. 日本語名詞文

日本語名詞文の基本形は「N1はN2です」である。主題の名詞「N1」について、「N2です」という解説を与えている文である。

日本語名詞文を特徴付ける一つのポイントは、日本語名詞自体の語彙量の多さである。日本語の名詞は他の品詞に比べて語彙数が圧倒的に多い。人名、地名などの固有名詞や、いろいろな物の名前、年月日、時間、値段、数字など、日常生活でよく使うことばの多くは名詞である。語種としては、和語、漢語、外来語、混種語と全てのものが含まれており、また、成り立ちも、動詞由来のものや形容詞由来のものがある。つまり、名詞は、多様性に富んだ語であり、それだけで多くの情報を交換することができると言える。

ここで、翻訳規則に関連すると思われる名詞の種類について、若干の説明を加える。

日本語名詞の中には「食事」、「勉強」、「遅れ」、「生まれ」のような、ある動作を表現する名詞がある。以下、こうした特徴を持つ名詞を動作性名詞(参考文献4)と呼ぶ。動作性名詞は、次の二種

類に分けられる。

一種類は「勉強」、「洗濯」、「入学」、「留学」などのように「する」を付けて「サ変動詞」となることが可能な名詞である。これらの名詞を「サ変名詞」と呼ぶ。このような名詞は目的や意志を表す。

例(1)この店は休業です。

例(2)試合は雨で延期です。

そして、「サ変名詞」の後ろに「中」を付け、全体として名詞述語にすることで、その動作が進行していること、またはある状態であることを表すことができる。

例(3)今は食事中です。

例(4)彼はアメリカに留学中です。

もう一種類は、「帰り」、「喜び」などのように、動詞の連用形を名詞に転成してできた名詞である。このような名詞を「動詞連用形名詞」と呼ぶ。これらは、すでに名詞として固定している感じがするが、まだ多少の動作性が残存していると言えるであろう。

例(5)私は1981の生まれです。

例(6)電車は20分遅れです。

また、日本語名詞の中には形式名詞と呼ばれるものがある。形式名詞とは、実質的な意味を失い、前にある言葉と一緒に使うことで全体を名詞化するという文法的な機能を果たす名詞である。「こと」、「ところ」、「もの」などは一般的な名詞でもあるが、形式名詞としてもよく使われる。

例(7)私の趣味は写真を撮ることです。

例(8)赤ちゃんは泣くものです。

3. 機械翻訳の困難性

現有の日中機械翻訳システムの名詞文翻訳上の問題点を洗い出すために、入力文数を増やして翻訳実験を行い、市販翻訳ソフトの問題点をまとめた。

3.1 入力文の選択

入力文はインターネットと「新潮文庫の100冊」で拾った名詞文1200文である。なお、選択した例文は以下の条件を満たす。

- ① うなぎ文ではない
- ② 文中に副詞句を含まない
- ③ 慣用表現を含まない単文である
- ④ 40音節以内のものである

3.2 翻訳ツール

翻訳ツールとして「蓬莱2008」を用いた。「蓬莱2008」は最新の日中、中日翻訳エンジンを搭載し、中国語と日本語の双方向翻訳に対応した翻訳ソフトである。翻訳実験では「蓬莱2008」で日本語を中国語に翻訳し、訳文を出力する。

3.3 現有翻訳システムに対する評価

評価基準は以下の3段階とした。

- (○) 正訳。文法が正しく、意味が理解できる。
- (△) 不自然な訳。文法に不自然なところがあるが、意味は理解できる。
- (×) 誤訳。理解できない。

評価値○、△、×の例をそれぞれ以下に示す。

○評価の例

入力文：妹は中学生です。

出力文：妹妹是中学生。

△評価の例

入力文：郵便局は市役所の北側です。

出力文：邮局是市政府的北側。

この例は対応文型が違うことによって、△評価となった。正訳は「邮局在市政府的北側。」である。

×評価の例

入力文：あの人は王さんです。

出力文：那些人是帝王。

この例文は単語の多義性によって、訳語の意味が分からなくなり、×評価となった。正訳は「那个人是小王。」である。

以上のような判断基準に基づき、1200文の例文の翻訳結果を分析し、評価値をつけて翻訳システムの翻訳精度を評価する。次に翻訳実験結果について分析する。

3.4 翻訳実験結果

3.1節で選択した1200文について、翻訳ツール「蓬莱2008」を用いて行った翻訳実験の結果を表1に示す。

表1 翻訳結果

評価値	件数	比率
○	452	37.7%
△	188	15.7%
×	560	46.6%

この表から見ると、○評価、つまり、正訳の比率は37.7%しかない。現有の翻訳ソフトの翻訳精度が不十分であり、精度向上の必要性があることが明らかとなった。

3.5 形態素解析実験

誤訳原因を、日本文の名詞という観点から明らかにするために、形態素解析ツール「茶筌」(奈良先端科学技術大学院大学松本研究室 version2.4.1)を用い、形態素解析実験を行う。形態素解析実験で、名詞を抽出し、名詞 N1 と N2 の意味的な関係を分析する。

4. 日本語名詞文と中国語の文型対応関係

前述のように、日本語では名詞文で表現されるものを、中国語では名詞文だけではなく、動詞文、「在」字文、「是」字文で表現することがある。これは日本語の名詞の語彙数が多く、両語の文法構造、表現形式が異なるために起こる現象であろう。次に、その訳し分けについて述べる。訳し分けの概略については、拙稿(参考文献7)で述べたが、今回はそれをさらに向上させた。

4.1 中国語「是」字文となるもの

中国語の「是」字文は「是」字を用いる文型である。日本語名詞文に対応する、最も一般的な対応文型である。日本語の「N1はN2です」と、中国語の「N1是N2」は、名詞と名詞が「は」あるいは「是」で結ばれるという共通点を持っている。

例(9)日本語 私は学生です。

中国語 我是学生。

4.2 中国語名詞文「N1N2」文となるもの

中国語名詞文「N1N2」文は「是」字文の「是」を省略して、名詞そのものが述語になる文型である。

日本語名詞文の後項名詞が、日付、曜日、時間、年齢、季節などの巡回性がある名詞の場合、中国語では名詞文「N1N2」文で対応する。その対応関係の詳細を4.5節の提案規則3,4にまとめる。

例(10)日本語 今日は5月5日です。

中国語 今天(是)5月5号。

4.3 中国語「在」字文となるもの

中国語の「在」字文はあるものがある場所に存在することを表現する文であり、場所名詞を伴う。場所名詞を用いる日本語名詞文の一部は、中国語では「在」字文で表現される。その対応関係の詳細を4.5節の提案規則6,8,9にまとめる。

なお、日本語の場所名詞の種類を以下のように分類した。分類にあたっては、(参考文献1)を参考にした。

- a 地名(日本、北京など)
- b 場所を表す意味を含んだ語(家、学校など)
- c 位置を表す語(方位詞、指示代名詞など)

例(11)日本語 日本語科は図書館の隣です。

中国語 日语科在图书馆的旁边。

4.4 中国語動詞文となるもの

中国語の動詞文は述語が動詞からなる文である。

日本語でサ変名詞、動詞連用形名詞、天候名詞など動作性が強い名詞を用いる名詞文の一部は、中国語では動詞文で対応する。その対応関係の詳細を4.5節の提案規則2,5,7,10,11にまとめる。

例(12)日本語 上野方面は次の駅で乗り換えです。

中国語 去上野的旅客在下站换乘。

4.5 提案規則

3.4節にまとめた翻訳結果を概観すると、○評価となった文の多くは日本語と中国語の文型の対応関係が単純な「是」字文であった。これに対し、翻訳結果が△評価あるいは×評価になった文には、

主に二つのパターンがある。一つは、本来「在」字文、動詞文、名詞文として訳されるべきものが、現有のソフトでは対応文型が「是」字文となるために誤訳となったものである。もう一つは、「是」字文として訳されるものの中で、日本語の名詞の意味判断が適切に行われなかったために誤訳となったものである。

今回行った翻訳実験結果を、日本語の名詞間の関係を視点として分析し、日本語名詞文と対応する中国語の文型について考察を加えた。これを元に、提案済みの翻訳規則に必要な修正を加えて新しい翻訳規則を作成した。その規則を表 2 にまとめる。

表 2 提案規則

No	N1	N2	対応文型		例	訳文
1	形式名詞	形式名詞	異なる文型		この本を買ったのは田中さんです.	买了这本书的人是田中.
					会議は今始まるところです.	会议现在正要开始.
2	名詞	サ変名詞+中	動詞文		委員たちは今食事中です.	委员们现在正在吃饭.
3	名詞	年齢・値段などの名詞	「N1N2」文		弟は二十歳です.	弟弟（是）二十岁.
4	相対的な時	日付, 時間, 曜日などの絶対的な時	「N1N2」文		明日は土曜日です.	明天（是）星期六.
					今日は5月5日です	今天（是）5月5号.
5	相対的な時	天候名詞	動詞文		今日は雨です.	今天下雨.
6	一般名詞	場所名詞 b, c	「在」字文		お手洗いは左側です.	洗手间在左边.
	人称代名詞				田中さんは事務所です.	田中在办公室.
7	場所名詞	サ変名詞	動詞文		ここは禁煙です.	这里禁烟.
		動詞連用形名詞			学校は休みです.	学校放假.
8	場所名詞 a	場所名詞 c	「在」字文		豊田市はここです.	丰田市在这里.
	場所名詞 b				愛知銀行はあそこです.	爱知银行在那里.
9	場所名詞 b	場所名詞 a	N1 の範囲 < N2 の範囲 「在」字文		私の学校は東京です.	我的学校在东京.
			N1 の範囲 >= N2 の範囲 「是」字文		日本の首都は東京です.	日本首都是东京.
10	場所名詞以外の名詞	サ変名詞	N2 を動詞に変換	変換後の文が正しい場合, 動詞文	試合は中止です.	比赛中止.
				変換後の文が正しくない場合, 動詞文	これはケーキの紹介です.	这个是蛋糕的介绍.
11	場所名詞以外の名詞	動詞連用形名詞	N2 を動詞に変換	変換後の文が正しい場合, 動詞文	試合は雨で取りやめです.	比赛因为下雨而终止.
				変換後の文が正しくない場合, 動詞文	それは異常の始まりです.	那是异常的开始.
12	以上の翻訳規則があてはまらない文		「是」字文		これは中国語の辞書です.	这个是中文词典.

5. 翻訳システム

機械翻訳の基本方式の主なものには、トランスファ方式、中間言語方式、用例ベース方式などがあるが、ここでは、トランスファ方式を採用する。

なお、用例ベース方式は、個人で対応できる作業量に配慮して採用を見送った。中間言語方式は、英語を中間言語とした翻訳実験を行ったが、望ましい結果が出なかったため、取り上げなかった。

トランスファ方式では、入力した日本語の文を形態素解析、構文解析をしたうえで、翻訳辞書、対応規則を参照しながら、日本語の構造、語彙を中国語に変換する。それから、訳文を調整して出力する。日本語の「N1 は N2 です」を中国語の「是」字文で対応する場合の流れを図 1 で示す。

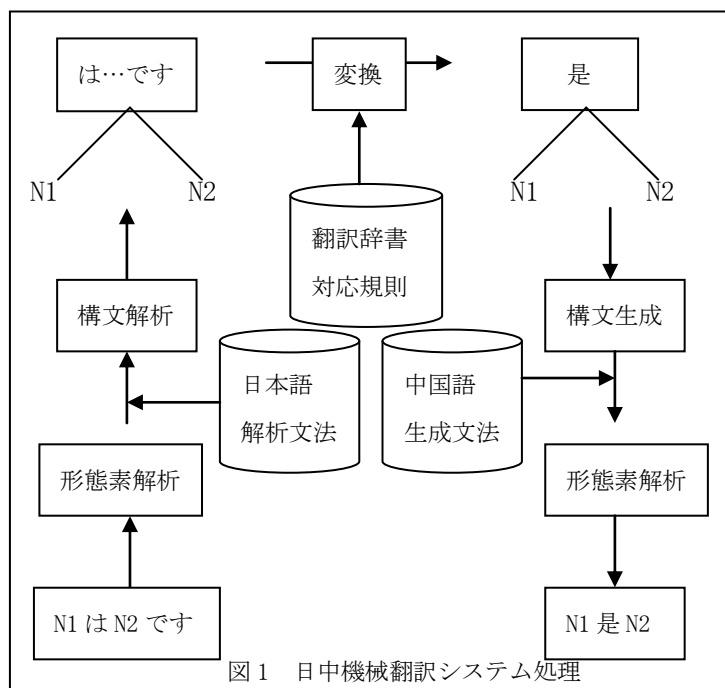


図1 日中機械翻訳システム処理

N1（私たち）：名詞-代名詞

N2（努力家）：名詞-一般

対応パターン：規則 12

対応文型：「是」字文

今回提案した規則によると、訳文は「是」字文になるが、実際には、この例文の「努力家」は強い動作意識を持つ名詞であるため、正訳は「我们很努力。」の動詞文である。現在の提案規則ではこのような対応規則がないため、正しく対応することができない。

対応できない事例を分析したところ、誤訳の原因の多くは日本語名詞文が同じ表現パターンであっても、中国語では対応文型が違うことにあった。この問題を解決するために、名詞の意味構造をより深く分析した上で、中国語との対応関係を見直す必要があると考える。

7. 終わりに

言語処理技術の発展により、機械翻訳の実現が可能になってきた。一方、現有の日中機械翻訳ソフトの精度が低いのも事実である。本研究では、日中機械翻訳における名詞文の翻訳精度を高めることを目標として、翻訳実験を重ね、名詞の意味分類を視点として誤訳の原因を考察した。これを元に、提案済みの翻訳規則に修正を加え、より精度の高い翻訳規則を提案した。今後は、今回提案した規則で対応できない文について検討するとともに、単文だけでなく、複文についての研究も行う予定である。

【参考文献】

- [1] 丸尾誠：中国語の場所詞について：モノ・トコロという観点から、言語文化論集, Vol. 25, No. 2, pp151-166 2004
- [2] 王軼謳ト, 池田尚志：日中機械翻訳における存在表現の翻訳処理について, 自然言語処理, Vol. 14, No. 5 pp. 65-105
- [3] 張建華：日中翻訳における表現構造に関する考察, <http://www2.kokken.go.jp/jalic/group5/98.8p77.pdf>
- [4] 庭三郎：現代日本語文法概説, 日本語文法研究会, 1991
- [5] 楊鵬, 村上仁一, 徳久雅人, 池原悟：日中機械翻訳に対する結合価パターン翻訳方式の応用, 言語処理学会第13回年次大会 C1-4, 2007
- [6] 出羽達也, 熊野明：中日・日中機械翻訳システム, 東芝レビューvol. 62No4, 2007
- [7] 陳麗, 山本いずみ：日中機械翻訳における「NはNです」文型の分析, 言語処理学会第14回年次大会 B2-2, 2008

6. 評価

6.1 評価方法

本研究では、提案手法による翻訳能力を評価するために、3.1節で選択した1200文を対象として、手作業によって、提案した翻訳規則に対して評価実験を行った。

評価過程を、以下の例を用いて説明する。

例：出口はあそこです。

N1（出口）：名詞-一般

N2（あそこ）：名詞-場所-代名詞

対応パターン：規則 6

対応文型：「在」字文

訳文出力：出口在那里。

例文に対して、N1, N2 の品詞を判断した上で、翻訳規則を用いて、対応文型を判断する。正確に対応文を導き出した場合、「提案規則は有効である」と判断する。

6.2 評価結果・考察

評価の結果、1200文中41文について、正確な対応文を導き出すことが出来ず、「提案規則は有効でない」という評価が出た。その原因は対応規則の不完全さにあると考えられる。

例：私の息子は大喜びです。

N1（息子）：名詞-一般

N2（喜び）：名詞-一般

対応パターン：規則 12

対応文型：「是」字文

今回提案した規則によると、この例文の訳文は「是」字文になる。ところが、この例の名詞「喜び」は形容詞の性質を持ったため、正訳は「我的儿子很高兴。」の形容詞文である。現在の提案手法では形容詞文で対応する規則がないため、正しく対応することができない。

例：私たちは努力家です。