

日中機械翻訳における連体修飾構造の翻訳について

† 朝暉 † † 浅井良信 † 宇野修一 † 池田尚志 †
Zhaohui Bu Yoshinobu Asai Syuichi Uno Takashi Ikeda

1 はじめに

日本語の連体修飾構造は一般名詞に対する連体修飾構造と形式名詞に対する連体修飾構造に分けることができる。一般名詞に対する連体修飾構造には、修飾語が直接に名詞を修飾する「XN」の形と、助詞「の」を介して名詞を修飾する「XのN」の形の2通りがある。中国語の一般名詞に対する連体修飾構造もそれと相似して、助詞「的」を介せずに「XN」と、助詞「的」を介して「X的N」の2種類の構成がある。しかし、「の」と「的」の対応にはずれがある。また日本語の形式名詞(「の/もの/こと」など。本稿では「の」について考察する)に対する連体修飾構造も中国語と相似しているが、「の」は中国語では「的」か「の」かに翻訳され、ずれがある。

我々は[6]で一般名詞に対する連体修飾表現の一部の日中翻訳規則とその翻訳処理について述べた。しかし、[6]ではシステムでの翻訳実験には至っておらず、形式名詞「の」の連体修飾構造の翻訳については考察していなかった。そこで、本論文では連体修飾構造全体を取り上げ、その構造における「の」と「的」の異同を分析した上で、日中機械翻訳のための翻訳規則を提案する。またこれらの規則について小規模な翻訳実験と手作業の評価を行い、問題点を検討する。

2. 一般名詞に対する連体修飾構造の翻訳

2.1 一般名詞連体修飾構造の「の」と「的」

日本語と中国語はそれぞれ膠着語と孤立語であり、基本語順はSOV,SVOと違いが大きい。連体修飾構造に関しては、中国語は日本語の「の」と同じく連体修飾の標識「的」をもち、語順も日本語の「X(の)N」と同じく「X(的)N」になるなど共通性が多く見られる。[他の言語、例えば、屈折語とされる英語や同じ孤立語のベトナム語では、名詞の所有格を使う、様々な前置詞を使う、修飾語と被修飾語の位置は日本語と逆になる。あるいは関係代名詞を使うなど異なりが大きい(my book, students in the classroom, the wife of the professor, the person whom I met)]。即ち、連体修飾構造には日中でそれぞれ次の2通りがある。

(日)(1) 修飾語+の+被修飾語

例: ①私の兄 ②先生への敬意 ③彼の本

(2) 修飾語+被修飾語

例: ④長い手紙 ⑤開けたドア ⑥この人

(中)(1) 修飾語+的+被修飾語

例: ③ 他的书 ⑤ 打开的门 ② 对老师的敬意

(2) 修飾語+被修飾語

例: ① 我哥哥 ④ 长信 ⑥ 这个人

† 岐阜大学工学部

‡ 中国広西大学外国語学部 530004

連体修飾語の標識としての「の」と「的」は対応している場合があるが(②→②、③→③と④→④)、ずれがある場合もある(①→①、⑤→⑤)。

「N1のN2」における「の」と「的」の異なりについて、文献[1]では「“的”にはシンタクスのレベルにおいて常に「省略」の可能性が存在するのに対して、「の」にはそれが考えられない…」と、また『「の」の「形式性強、意味性弱」に対し、中国語の“的”は、…「形式性弱、意味性強」であると言える。』と指摘している。

日本語では名詞が名詞を修飾する場合、すべて「の」を介し、用言が名詞を修飾する場合は「の」を介さない。文法規則としてははっきりしている。しかし、中国語の連体修飾語の「的」の使用は、統語的、意味的、習慣的、音節などの影響を受けて総合的に決定され、日本語のような単純な原則はない。例えば、名詞が名詞を修飾する場合は「的」の使用と不使用は意味的な影響が多く見られるが、習慣用法や、構語法とも関連がある。形容詞が名詞を修飾する場合は音節や、文脈などと絡まる。動詞が名詞を修飾する場合、ほとんど「的」を使う。それは統語の規定の作用が大きい。原則からはずれた特殊な用例も見られる。

2.2 一般名詞に対する連体修飾構造の翻訳

2.1節の分析から、一般名詞に対する連体修飾構造の日中翻訳では、「的」の訳出は中心的な問題であることが分かる。我々は機械翻訳の便利を考えて、日本語の連体修飾構造を次の6種類に分け、翻訳規則を作成した。分類と訳し分け条件は文献[3][4][5][8]などの論述と例文を参考にして、連体修飾語と中心語の統語的、意味的な属性を用いてまとめた(意味属性としては主に文献[9]を参考にした)。我々が開発している日中機械翻訳システム jaw/Chinese[7]でのこれらの規則の実装についても簡単に説明する(以下 jaw/Chineseをj/Cと略す)。

(1)N1+の+N2 (N1=名詞、代名詞、数量詞)

例: 私の妹→我妹妹 一匹の猫→一只猫
花子のカバン→花子的书包

このタイプの場合、日本語ではいつも「の」を使うのに対して、中国語は場合によって「N1N2」あるいは「N1的N2」となる。我々はN1とN2の意味的、統語的な属性などを用いて「N1N2」に翻訳する場合の規則をまとめた(表1)。これ以外の場合には「N1的N2」に翻訳する。

ただし、上記以外に以下のような特殊な場合もある。

(1-1) 「的」の代わりに古語の「之」を使う場合。

例: 労使の間→雇佣之间

(1-2) 「N1+の+N2」構造を一つの語彙として訳す。

例: 農家の人→农民

j/Cでの実装: 「の」をキーワードとする2種類のパターンを作成する。「N1N2」に翻訳する場合はそれぞれ表1の属性条件を書き込む。「N1的N2」に翻訳する場合は、デ

フォルトのパターンとして任意の属性に適合する意味条件を書き込んでおく。(1-1)-(1-2)のような特殊な場合はそのまま辞書登録する。

(2) N1+格助詞/格助詞相当語+の+N2

- 例：a 母親としての資格→作为母亲的资格
b 日本での生活→在日本的生活

このタイプは中国語では基本的に「的」が必要である。ただし、格助詞によっては、中国語の助詞への訳しわけが必要になる場合がある。我々は曖昧性が多い「で」、「へ」の基本的な訳しわけ規則を作成したが未だ十分なものではない。これについて別の機会に述べる。

翻訳規則：N1+格助詞/格助詞相当語+の+N2

→介詞+N1+的+N2

j/Cでの実装：「での」のような「格助詞+の」をキーワードとするパターンを作る。中国語への訳しわけがある場合は訳しわけの属性条件をパターン中のN1,N2の属性条件欄に書き込む。

(3) Adj+N (Adj=形容詞或いは形容動詞)

- 例：a 悪い人→坏人
b 綺麗な部屋→干净的房间

このタイプは日本語では「の」を使わないが、中国語では「的」を使う場合と使わない場合がある。その使用に関しては言語上で習慣性から、語義関係から、音節から様々な論述が見られる。我々は音節からの分析に基づき下記のように規則化した。文字数は本来音節数とすべきであるが、機械処理の便宜上文字数とした。

翻訳規則：① Adj+N→Adj+N

(中国語のAdjが一文字の場合)

② Adj+N→Adj+的+N

(中国語のAdjが2文字以上の場合、あるいは形容詞が直前で副詞に修飾されている場合)

j/Cでの実装：中国語の表現構造から線状化する際に、プログラムで処理する。音節数を文字数として、上述の規則をプログラム化する。

(4) V+N

- 例：a ヌードルを食べる音→吃面条的声音
b 母が書いた手紙→妈妈写的信

この場合日本語では「の」を使わないが、中国語では基本的に「的」が必要である。

翻訳規則：V+N→V+的+N

j/Cでの実装：2つのタイプに分けて処理する。

① NがVと内関係の場合(「母が書く手紙」)はVをキーワードとする基本パターン(母が手紙を書く)から変換してパターン照合する。その際に「的」が必要という情報をプログラムで書いておく。

② 「食べる音」のような外関係の場合は、被修飾名詞の「音」をキーワードとするパターンとの照合で処理するが、そのパターンの中に「的」が必要という情報を書いておく。

(5) V+助詞+の

- 例：a 見てからの決定→看以后的决定
b 読んだ後の感想→读了以后的感想

この場合日本語では「の」を、中国語では「的」を使う。

翻訳規則：V+助詞+の→V+助詞+的+N

j/Cで未だ実装していないが、「V/てからの/N」のように解析し、機能語「てからの」をキーワードとするパターンの翻訳規則として記述する予定である。

(6) 連体詞/連体形容詞+N

このタイプは次の①と②に分ける。このタイプの連体修飾語は日本語では助詞「の」がないが、中国語では①タイプは「的」を使わずに、代わりにNを数える量詞が必要になる。②タイプは「的」を使う。

① 連体詞+N

- 例：a この人→这个人
b 「ある人」→某个人

翻訳規則：連体詞+N→連体詞+量詞+N

② 連体形容詞+N

- 例：c そのような数字→那样的数字
d いわゆる異星人→所谓的异星人

翻訳規則：連体形容詞+的+N→連体形容詞+的+N

j/Cでの実装：①の場合、中国語ではNを表す量詞が必要であるため、「N」をキーワードとする個々のパターンの中に、その「N」に対して用いる量詞を記述しておく。

②の場合、「このような」などの連体形容詞をキーワードとするパターンの中に「的」を用いるという情報を記述しておく

3. 形式名詞に対する連体修飾構造の翻訳

3.1 形式名詞の「の」と「的」

ここで言う「形式名詞」の「の」は「他の語について、ある意味を加え、全体として体言と同等の機能をもつ単位を作る助詞」[2]であり、準体助詞とも呼ばれる(ここでは「の」以外の形式名詞については議論しない)。例えば、

- (1) 東京へ行ったのが一年生だ。
→ 去东京的是一年级学生。
(2) コーヒーは熱いのがいい。
→ 咖啡热的好。
(3) 私は友達の李さんが来るのを待っている。
→ 我等着朋友小李来。(の=φ)
(4) うまく作ったのを選ぶ。
→ 选做得好的。
(5) この本は私のだ。
→ 这本书是我的

形式名詞の「の」を用いる表現も文全体の構造としては中国語と相似しており、多くの場合は「の」を同じ名詞化の作用を持つ「的」で置き換えればよい[例(1)(2)(4)(5)]。しかし「φ」に翻訳される場合もある[例(3)]。

構文上から見ると、日本語の従属節はすべて「の」によって体言の資格をもたされなければならないのに対して、中国語では従属節は「的」によって名詞化する場合もあるし、「的」を使わずに動詞節そのままで名詞節として扱われる場合もある。すなわち形式名詞の「の」も中国語では

「的」か「 ϕ 」に訳すという2つの可能性が出てくる。しかし単に統語的な面上からだけでは形式名詞の「の」がどんな場合に「的」に、どんな場合に「 ϕ 」になるかは説明しきれない。例えば(3)と(4)の従属節は中国語では同じ目的語であるが、(3)は「的」を使用せずに(4)は「的」を使う。それは「の」に託された下記の意味用法上の違いから考えるべきである。

意味用法から形式名詞「の」は主に次の2種類に分けられる。

a) 用言の表す事柄をそのまま体言化する [例(3)]

b) 用言の表す事柄によって限定された「もの」を表す。

[例(1)(2)(4)]

[(5)は2.1節の連体修飾構造の被修飾部が省略された類であり、「の」は体言につき、「…もの」の意を表す。]

b)の場合の「の」は「もの」を表し、中国語では「的」に、a)の場合は「事柄」を表し、中国語では「 ϕ 」に翻訳される。例えば(1)にある「の」は「学生」であり、「人」である。(2)は「コーヒー」であり、「もの」である。(4)は確実には分からないが、ある「もの」である。中国語ではすべて「的」に訳す。(3)は「人を待つ」事柄であり、「 ϕ 」に訳される。

3.2 形式名詞「の」の翻訳規則

3.1の分析から、形式名詞の連体修飾構造を中国語へ翻訳する際、「の」の翻訳は重要な課題であることが分かる。我々は文献[2][4][10]などの例文を参考にして、日本語文の構成特徴、[の]に後接する助詞、従属節と主節の述語の統語的、意味的な属性などから形式名詞「の」の翻訳規則を作成した。規則は11個あるが表2でその一部を示す。

4. 翻訳実験と問題点

4.1 一般名詞に対する連体修飾構造の翻訳実験

このタイプの連体修飾規則をjaw/Chineseに実装して小規模な実験を行った(表3)。実験に用いた例文は規則作成した際に参考した例文とは別に、「英作の基本文例」(池田書店, 1989)から抽出した。

表3 jaw/Chineseにおける連体修飾語の実験

実験文	連体修飾語の数	正訳	誤訳	正訳率
186	85	81	4	95%

4.2 形式名詞「の」の翻訳評価

表2の形式名詞の規則を用いて手作業による評価を行った。評価に用いた例文は規則作成した際に参考した例文とは別に、4.1と同様の英作の基本文例と、「講談社和英辞典」(講談社, 1976)から抽出した。評価結果は表4で示す。

表4 形式名詞「の」の評価

評価文	正訳文	誤訳文	正訳率
124	116	8	94%

4.3 問題考察

[A] 一般名詞に対する連体修飾構造の翻訳実験の問題

(i) タイプ(1)の連体修飾語に関して、現規則で対応できない例があった。

① 利害の関係が労使の間に存在する。

現訳：利害的关系存在于雇佣之间。

正訳：利害 ϕ 关系存在于雇佣之间。

分析：我々の規則によって、「の」は「的」に翻訳されているが「NIN2」に翻訳するべきである。日本語では「利害の関係」という連体修飾構造の表現のほか、「利害関係」という複合名詞もあるが、中国語では「利害關係」という複合名詞の形しか取れない。即ち、連体修飾構造→複合名詞。これは造語法の違いに関連する問題であり、これに関する考察は今後の課題である。

(ii) 多層修飾の場合、語順や、「的」の使用の変化がある場合があるが、現規則ではまだ対応できていない。

② 蝶は甘い香りの蜜を吸う。

現訳：蝴蝶吸甜香味的蜂蜜。

正訳：蝴蝶吸香甜的蜂蜜。

③ 祖父は果物の出荷の分別作業で疲れた。

現訳：祖父因为水果的出货的分类作业感到疲劳。

正訳：祖父因为水果 ϕ 出货的分类作业感到疲劳。

分析：2例とも多層の連体修飾構造である。②は、中国語では「香りの甘い蜜」という語順になっている。誤りの原因としては、日本語では「甘い香りの蜜」という「形容詞+名詞+の」の連体修飾語であるのに対して、中国語では「香甜」を一つの複合形容詞的のものとして見られ、「形容詞+的」という一層の連体修飾語になっていると関係がある。また両言語の品詞上のずれとも関わっている(日本語では「香り」=名詞、中国語では「香」=形容詞)。

③では「果物の出荷」、「出荷の分別作業」というような一層の連体修飾構造ならば、両方も「的」を使う。しかし、「果物の出荷の分別作業」の多重修飾になると、2番の「的」のみを使うのが自然である。このように、多重連体修飾構造に関する考察が必要になってくる。

(iii) 規則の意味属性の付与の問題

例えば表1の⑧番の「材料」という意味属性は我々の辞書にないため、代わりに「資材/物質」という属性を付与して翻訳を行っている。そうすると、対応しきれない場合が出る(「材料」イコール「資材/物質」ではないため)。

④-1 石のテーブル→石头 ϕ 桌子

④-2 石の重さ→石头的重量

人間には、④-1の「石」をテーブルの「材料」、④-2の「石」を普通の「物質」として理解でき、訳しわけが可能である。しかし、機械辞書では2例の「石」とも同じ「物質」の属性しか付与できないため、④-2を④-1と同様に訳される。「材料」というような社会知識や文脈の理解に関わる属性を機械辞書での付与は困難である。

[B] 形式名詞の問題

(i) 6-2番の規則に関する問題

⑤ 一番印象に残ったのは主役の演技だった。

現訳：最难忘 ϕ 是主角的演技。

正訳：最难忘的是主角的演技。

分析：これは表の6-2番の規則に適用して(演技=抽象名詞)「の」を「 ϕ 」に誤訳した。これを解決するためには「N」の意味属性に対する細分類が課題として残っている。

(ii) 構造のずれの問題

⑥ 彼女は家柄のよいのを自慢している。

現訳：她为家庭出身好的感到骄傲。

正訳：她为家庭出身好感到骄傲。

分析：これは3番の規則の条件と照合して(よい=形容詞)「的」に誤訳した。日本語では「の」に前接する節は主節の動詞「自慢する」の目的語になる。しかし中国語ではそれは介詞(前置詞)「为」で導く連用修飾節になる。このように構造上の変化は形式名詞「的」の使用に影響がある。日中の形式名詞文の構造は基本的に相似しているが、まれにこのような構造のずれが出ることもある。

6. まとめ

本論文は日中両言語における連体修飾構造及び形式名詞文における「の」と「的」の異同を分析し、機械翻訳の規則としてまとめた。小規模であるが、連体修飾構造に対しての機械での実験では約95%、形式名詞に対する手作業の翻訳評価においては、約94%の正訳率を得た。今後はさらに多くの例文を用いて実験を行い、規則をより充実していく予定である。形式名詞「の」の規則のシステムへの組み込みも当面の課題である。また、一般名詞に対する多層連体修飾構造の翻訳に関する考察も行う予定である。

謝辞：この研究の一部は日本学術振興会外国人特別研究員事業の援助によって実施したものである。

参考文献：

- [1] 杉村博文. “的”と「の」, 『中国語』(No.368). 1990.
- [2] 松村明編. 『日本文法大辞典』, 明治書院. 1971年.
- [3] 房玉清. 『实用汉语语法』, 北京大学出版社 2001.
- [4] 水野義道. 日本語の「の」と中国語の「的」. 日本語学, VOL.12, 10, 1993.
- [5] 劉月華, 潘文娛, 故韓. 『現代中国語文法総覧』, くろしお出版, 1996.
- [6] 宇野修一, 卜朝暉, 今井啓允, 池田尚志. 中機械翻訳システム jaw/Chinese における連体修飾表現、場所表現などの処理, 言語処理学会第10回年次大会発表論文集 pp.640-643.
- [7] 今井啓允等. 日本語からアジア諸言語への機械翻訳の試み, 情報処理学会第65回全国大会講演論文集, pp.5-363~5-366, 2003
- [8] 森田良行. 『日本語文法の発想』, ひつじ書房. 2002年.
- [9] 池原, 宮崎, 白井 他. 日本語彙大系, 岩波書店. 1997.
- [10] 名柄, 広田, 中西. 外国人のための日本語例文問題シリーズ2 形式名詞. 荒竹出版. 昭和62年.

表1 「N1+の+N2」を「N1+N2」に訳す場合の判定条件

判定条件	翻訳規則	例文
①数量詞+の+名詞(N1=分数→の=之 N1=重ね型の数量詞/数値→の=的)	「N1+N2」	一匹の猫→一只猫(三分の一→三分之一 3メートルのリボン→3米(长)的彩带)
②性別名詞+の+名詞	「N1+N2」	雄の兔→雄兔 女性の研究者→女性研究者
③(中)単音節人称代名詞+の+親族関係名詞	「N1+N2」	私の妹→我妹妹
④職名+の+人名名詞(称呼名詞)	「N1+N2」	校長の田中先生→校长田中先生
⑤時間名詞+の+時間名詞(特殊の場合がある)	「N1+N2」	今日の午後→今天下午(来年の今日→明年的今天)
⑥名詞+の+場(上下/左右/前後/方向等)名詞	「N1+N2」	運動場の北側→运动场北边
⑦学科名詞+の+名詞	「N1+N2」	体育の授業(先生)→体育课(老师)
⑧材料(資材/物質)名詞+の+名詞	「N1+N2」	石のテーブル→石头桌子
⑨名詞1+の+名詞1の同格名詞	「N1+N2」	同僚の花子→同事花子

表2 形式名詞「の」の翻訳規則(一部)

番号	パターン構成	従節述語の属性	主節述語の属性	【の】の訳	例文
3	Adjのは(が/を)P	形容(動)詞	述語	的	(1)安いのがいい。→ 便宜的的好。
5	VのをP	動詞	用言	φ	(2)先生が飲食店でコーヒーを飲むのを見かけた→看见老师在饮食点喝咖啡。
6	1 P1のは(が)Nだ	用言	N=具体名詞	的	(3)メガネをかけているのは私の姉だ → 戴眼镜的是我的姐姐。
	2 P1のは(が)Nだ	用言	N=抽象名詞	φ	(4)そのような人に出会ったのは不幸の 始めだ→遇见那样的人是不幸的开始。
8	1 P1(V)のはAdj	目的語を伴わない他動詞	形容詞	的	(5)私がさっき読んだのは一番新しい → 我刚才读的最近。
	2 P1(V)のはAdj	自動詞/目的語を伴う他動詞	形容詞	φ	(6)言外の意味を汲み取るのは難しかった→(要)理解言外之意很难。
9	P1(V)のはAdj	動詞	形容動詞だ形	φ	(7)手紙を赤いペンで書くのは失礼だ →用红笔写信很失礼。