

日中機械翻訳システム jaw/Chinese における連体修飾表現、場所表現等の処理

宇野修一 ト朝暉 今井啓允 池田尚志

岐阜大学工学部

1 はじめに

我々の研究室ではパターン変換方式の日中機械翻訳システム jaw/Chinese の開発を行い、日本語から中国語への翻訳規則の記述を進めている。本稿では jaw/Chinese における連体修飾表現、場所表現の扱いについて述べる。

日本語の連体修飾表現は「N の N」や「V する N」というような形をとるが、中国語では一般に「A 的 B」という形式で表現する。しかし常に「的」が使用されるわけではなく、「A B」のように表現する場合もある。このため精度の高い機械翻訳を行うには連体修飾表現の翻訳時に訳語に「的」を用いるかを判定する機能が必要となる。

名詞には単独で場所として扱うことのできる名詞と単独では場所として扱うことのできない名詞とがあり、名詞をどちらの種類に分類するかは言語によって異なる。この分類の違いは訳文中の方位詞の有無や前置詞の選択に影響するため、場所表現を含む機械翻訳では名詞をそのまま翻訳するだけでは訳文としては不自然となる場合が多く、場所表現に関する翻訳の際に入力文中には存在しない表現要素を訳文中に自動的に補完する機能が必要となる。

2 機械翻訳エンジン jaw

jaw は VC++ で記述されたパターン変換方式の機械翻訳エンジンである。パターン変換方式はトランスファー方式の一種であり、入力文の表層の表現構造と深層の意味の両方を考慮して翻訳を行う。

jaw は入力文の木構造 (Input Tree・IT) の作成、翻訳規則の木構造 (Transfer Tree・TT) への変換、目的言語の表現構造木 (Expression Tree・ET) の生成、ET の各オブジェクトの線状化 (目的言語文の生成) の 4 つの処理で原言語を目的言語に翻訳する。jaw/Chinese は目的言語を中国語とする日中機械翻訳システムである。

IT の作成には本研究室で開発中の日本語解析システム IBUKI が用いられる。IBUKI は入力文の文節解析・構文解析を行うことで IT を作成する。TT への変換は IT と表現パターン変換辞書との照合によって行われる。翻訳規則は日本語入力文の係り受け木構造とそれに対応する目的言語の表現構造の対で、目的言語の表現構造を作

り出すプログラム (dll) として格納されている。ET の生成は TT の各翻訳規則関数を実行することで行われる。ET のノードはクラスオブジェクトで表現され、各オブジェクトのメソッドである線状化関数を実行することで木構造が目的言語の文となり、目的言語への翻訳が完了する。

図 1 に jaw の基本的な処理の流れを示す。

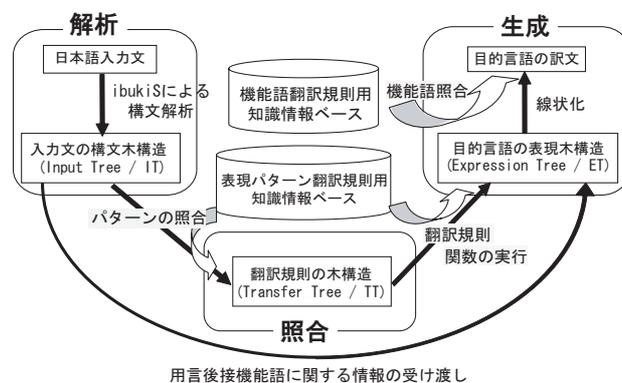


図 1: jaw の基本的な処理の流れ

3 jaw/Chinese における連体修飾表現の処理

3.1 中国語の連体修飾表現の特徴

日本語には名詞、形容詞、形容動詞、動詞、連体詞などを連体修飾語とする連体修飾表現が存在し、その品詞によって「『の』を用いて修飾する」「連体形で修飾する」など修飾方法が異なる。それに対し中国語の連体修飾表現は基本的に連体修飾語の品詞にかかわらず「A 的 B」という形を取ることが多い。しかし一部、連体修飾語と被連体修飾語の意味関係、音節数などによって「的」が省略され、「A B」の形になるものも存在する。したがって精度の高い日中翻訳を行うためには中国語の連体修飾表現における「的」の使用条件に関する規則が必要となる。

3.2 jaw/Chinese における処理方法

中国語の連体修飾表現における「的」の使用状況を分析し、表 1 のようにまとめた。表 1 の規則のうち品詞に関する条件と名詞単独の意味に関する条件は jaw におけ

表 1: 中国語の連体修飾表現における「的」の使用状況

番号	条件	的	例
1	動詞 + 名詞	的	熊が歩く道 熊走的路
2	単音節形容詞 + 名詞		小さな村 小村庄
3	多音節形容詞 + 名詞	的	美しい白鳥 美丽的天鹅
4	単音節人称代名詞 + の + 人称代名詞		私の妹 我妹妹
5	職業名 / 役職名 + の + 名前		校長の佐藤さん 佐藤校长
6	量詞 + の + 名詞		2匹の猫 二只猫
7	時間名詞 + の + 時間名詞		今年の秋 今年秋天
8	名詞 + の + 同格の名詞		同僚の佐藤さん 同事佐藤
9	名詞 + の + 方向 / 位置を表す名詞		学校の北側 学校北边
10	その他の名詞 + の + その他の名詞	的	私の机 我的桌子

表 2: 各規則における「A の B」処理用変数の値

番号	条件	変数 m_N_de の値	変数 m_TekiType の値
4	単音節人称代名詞 + の + 人称代名詞		SyllableCheck
5	職業名 / 役職名 + の + 名前		PositionChange
6	量詞 + の + 名詞		NounQuantifier
7	時間名詞 + の + 時間名詞		Normal
9	名詞 + の + 方向 / 位置を表す名詞		
10	その他の名詞 + の + その他の名詞	的	

る通常のパターンの照合処理で判別が可能だが、それに該当しない以下の6つの規則では特殊な処理が必要となる。

- 規則 2 音節数に対する判定
- 規則 3 音節数に対する判定
- 規則 4 音節数に対する判定
- 規則 5 語順の入れ替え
- 規則 6 量詞に対する処理
- 規則 8 同格に対する意味の判定

規則 2~4 の音節数に対する判定は訳語の文字数から行うことができる。jaw/Chinese では訳語が一文字のものは単音節、二文字以上のものは多音節として扱っている。語順の入れ替え、量詞に対する特殊処理が必要な規則 5~6 は専用のフラグから判別し、必要な処理を行う。規則 8 の同格に対する意味の判定には文脈の解析を必要とするが jaw は文脈の判定に対応していないため、現在のところ規則 8 については判定を行うことができない。

jaw/Chinese では図 2 のような構造から連体修飾表現の線状化を行っている。表 1 より、連体修飾語が動詞である場合 (「m_verb != NULL」の場合) は常に的を用いて線状化が行われ、連体修飾語が形容詞、形容動詞である場合 (「m_adjective != NULL」の場合) は常に音節数をチェックしてから線状化が行われる。「A の B」の形

を取る名詞による連体修飾表現では規則によってチェックする項目が異なるため、それらを判別する変数を用意して対応した。m_N_de はその規則を用いた場合の訳語中の「的」の有無を、m_TekiType はその規則で必要となる特殊処理の内容を表す変数である。各規則における変数 m_N_de と m_TekiType の値は表 2 に示す。

以上の処理を追加することにより、中国語の連体修飾表現に対するより精度の高い翻訳が可能となった。

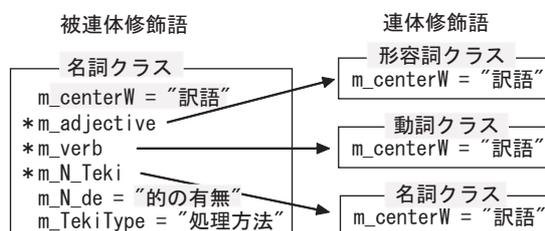


図 2: 連体修飾表現の構造

表 3: 日本語、中国語、英語における場所詞の分類の違い

例文	修正例			メモ
	日本語	中国語	英語	
彼はドアにいる	ドアのところ	门 这儿	near the door	3言語とも「ドア」は場所詞としない
リンゴは冷蔵庫にある		冰箱 里	in the icebox	中国語と英語では「冷蔵庫の中」と表現する
リンゴを冷蔵庫に入れる		冰箱 里	in the icebox	中国語と英語では「冷蔵庫の中」と表現する
絵が壁にかかっている		壁 上	on the wall	中国語と英語では「壁の上」と表現する
絵が壁の上にかかっている	壁(の上)			日本語では通常「壁の上」とは表現しない
彼は学校にいる			in the school	中国語では建造物の多くは場所詞に分類される
彼は図書館にいる			in the library	中国語では建造物の多くは場所詞に分類される
彼は日本にいる			in Japan	中国語では国名などは場所詞に分類される
私はその話を彼から聞いた		他 那里		中国語では「彼のところから」と表現する
彼はここにバッグを置いた				英語では「here」を用いるので前置詞は不要
彼はそこにバッグを置いた				英語では「there」を用いるで前置詞は不要

4 jaw/Chinese における場所表現の処理

4.1 言語毎の場所表現の特徴

日本語では以下の2つの例文

- 彼は食堂にいる。
- 彼は机にいる。

のうち、前者は自然な表現であるが後者は不自然な表現である。後者を不自然であると感じるのは日本語の「N1がN2にいる」という表現パターンではN2に用いることができる名詞が「場所として扱うことができる名詞」のみであるのに対し、例文中でN2に該当する名詞「机」が場所として扱えないことが原因である。

このように名詞には場所として扱うことのできる名詞と場所として扱うことのできない名詞の2種類が存在し、その分類は言語によって異なる。中国語では場所として扱うことのできる名詞を「場所詞」と呼ぶ。先ほどの例から日本語では「食堂」は場所詞、「机」は一般名詞であることがわかる。(本論文では便宜上、場所として扱うことのできる名詞は中国語以外の文中でも場所詞と呼ぶこととする。)日本語の名詞における場所詞の占める割合は中国語のそれよりも高く、日本語では多くの名詞を場所として使用することができる。また英語では場所詞に分類される名詞は中国語よりさらに少数で、「here」「there」「home」「abroad」など極一部の名詞性の副詞のみである。

場所詞ではない名詞を場所として使用したい場合の表現の変更方法も言語ごとに異なっている。例えば日本語であれば「～の中」「～の上」などの語を名詞の直後に、英語であれば「in」「on」「at」などの前置詞を名詞の直前に置くことで名詞を場所として使用可能にする。中国語では日本語と同様に「～の中」「～の上」などの意味の語を名詞の直後に置くことで名詞を場所詞化すること

ができる。中国語では「～の中」「～の上」などの意味の語を方位詞と呼ぶ。(本論文では便宜上、日本語における「～の中」「～の上」などの語も方位詞と呼ぶこととする。)

日本語、中国語、英語における場所詞の分類の違いを原因とする表現の違いとその修正例を表3に示す。

言語間の場所詞の分類の違いは機械翻訳において場所詞の翻訳を複雑にする。なぜなら表3からもわかるように原言語である日本語で場所詞に分類される名詞が目的言語では場所詞ではなかった場合、目的言語で「場所ではない名詞が場所として使用されている」という現象が起こるためである。つまり場所表現を含む文の機械翻訳では入力文での場所詞が目的言語において場所詞に分類されるかを識別し、場所詞でなかった場合は入力文中では表現されていない方位詞、前置詞を自動的に補完して出力するという処理が必要となる。

4.2 方位詞の種類と使用機会の依存関係

4.2.1 方位詞の種類と依存関係

表3で示したように名詞「冷蔵庫」は日本語では場所詞であるが中国語・英語では場所詞ではない。よって日本語入力文が「りんごは冷蔵庫にある。」であれば、中国語・英語の訳文では日本語入力文中に表現されていない方位詞、前置詞をそれぞれ補完して翻訳する必要がある。

名詞「冷蔵庫」に対し使用可能な方位詞は文法的には「～の中」「～の上」「～の下」「～のところ」「～裏」など様々な種類がある。この中で「りんごは冷蔵庫にある。」の訳文に対して補完する語は「～の中」の意味の方位詞「里」、前置詞「in」であると考えるのが妥当である。なぜなら日本語で「冷蔵庫の上にある。」や「冷蔵庫の下にある。」と言いたい場合に「～の上」「～の下」という

表 4: 方位詞の補完が不要な場合の例

補完の不要な入力文	誤った補完の例	補完が不要な理由
電車で通学する。	電車の中で通学する。	場所を表現するパターンではない
机の下に隠れる。	机の下の上に隠れる。	入力文中で方位詞が使用されている
彼は家にいる。	彼は家の中にいる。	場所を表す名詞が場所詞である

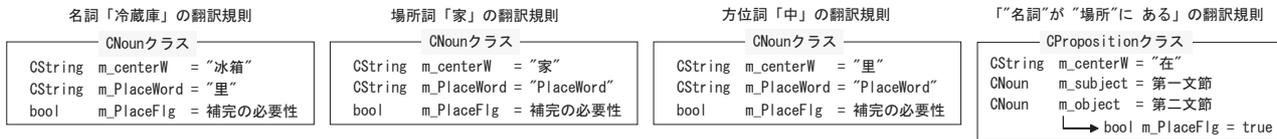


図 3: 方位詞の補完に関する翻訳規則の例

方位詞が文中から省略されるとは考えられないためである。また同じ「ある」という動詞であっても「テーブルにある。」という表現であれば「テーブルの中」ではなく「テーブルの上」だと考えられる。つまり日本語の入力文中で単に「冷蔵庫に」とだけ言っている場合は「冷蔵庫の中」、「テーブルに」とだけ言っている場合は「テーブルの上」の意味であるという関係が得られる。

これらのことから補完に使用する方位詞の種類は動詞パターンに依存するのではなく、補完が必要な名詞に依存していることがわかる。名詞と補完する方位詞、前置詞の関係の一部を表 5 に示す。

表 5: 名詞と補完する方位詞、前置詞の関係

名詞	方位詞	前置詞	例
電車	里	in	電車の中であった。
机	上	on	机の上で食べる
彼女	这儿		彼女のところから聞く
壁	上	on	壁の上にかかっている

4.2.2 方位詞の使用機会の依存関係

原言語と目的言語で分類の異なる「冷蔵庫」のような名詞であっても常に方位詞の補完が必要なわけではない。補完が必要なのはあくまで「冷蔵庫」が場所として用いられた場合のみであり、例えば「私は冷蔵庫を買った。」というような文の場合は当然「私は冷蔵庫の中を買った。」などとする必要は無い。このことから方位詞の補完が必要となる機会は動詞パターンに依存すると考えられる。ただし場所を表現している動詞パターンであっても「入力文中で方位詞が既に使用されている場合」と「場所の位置で使用されている名詞が中国語においても場所詞である場合」は方位詞の補完を行う必要は無い。方位詞の補完が不要である場合の例を表 4 に示す。

4.3 jaw/Chinese における場所表現の処理方法

jaw/Chinese では 4.2 節で述べたように方位詞の種類は名詞に、方位詞を補完する機会は動詞パターンに依存

するとの判断から、名詞クラスに「その名詞を場所として使用する際に頻繁に省略される方位詞」をセットする変数 m_PlaceWord と「線状化の際に方位詞の補完が必要か」を判定するフラグ変数 m_PlaceFlg を用意し、場所表現を含む動詞パターンの翻訳規則から方位詞を補完する必要のあるオブジェクトのフラグ変数 m_PlaceFlg を ON にすることで方位詞の補完処理を行う。

表 4 にあるような「方位詞が既に使用されている」「場所を表す名詞が場所詞である」などの理由により方位詞の補完が必要ない場合の判定は、方位詞と場所詞の翻訳規則で変数 m_PlaceWord に特殊値を指定することによって行う。つまり名詞クラスの線状化ではフラグ変数 m_PlaceFlg が ON であっても補完する方位詞を示す変数 m_PlaceWord の値が特殊値であれば線状化は行わないということになる。方位詞の補完に関連する翻訳規則の例を図 3 に示す。

5 おわりに

中国語の連体修飾表現における「的」の使用法、中国語の場所表現における場所詞と方位詞の使用法について分析し、それらに関する規則を jaw/Chinese に実装した。今回作成した連体修飾表現に関する規則は連体修飾語が 1 つのみの場合のものであり、「彼の学校の先生」のような「A の B の C」の形の表現など 2 つ以上の連体修飾語を使用する場合、「彼に関する話題」など訳語に介詞が必要になる場合などには対応していない。これらについては今後の課題としたい。

参考文献

- [1] 今井 啓允:『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 (2003)
- [2] 瀬戸口 律子:『中国語の文法』(語研、2003)
- [3] 相原 茂, 石田 知子, 戸沼 市子:『Why?にこたえるはじめての中国語の文法書』同学社 (1996)