

意味属性による中国語補助語の推定アルゴリズム†

任 福 継^{††} 宮 永 喜 一^{††} 栃 内 香 次^{††}

中国語では文を構成する単語間のあらゆる文法関係は“詞”(独立品詞)の語順と“辞”(補助語)との二つの手段だけで表されている。機械翻訳の規則を作製する際に考慮されなければならないのも、まさにこの点である。われわれは、日中両言語の特徴に基づいて、日中機械翻訳システムの研究開発を進めている。日中機械翻訳を行うためには、①独立品詞の語順の確定、②補助語とくに日本語格助詞に対応する補助語の推定が必要と考えられる。われわれはすでに①の研究を発表しており、本論文では②について述べる。一般に、日本語の格助詞に対応する中国語補助語は一对多であるので、複数補助語候補から正しい補助語を推定することは日中機械翻訳を行うとき解決しなければならない問題である。本論文では、まず、教科書、文献など約 12,000 文から格助詞およびその関連情報を抽出し、この情報を分析してから、意味属性による日本語格助詞に対応する中国語補助語を推定する手法を提案する。そして、格助詞の多義性を解消する情報としている日本語格助詞と中国語補助語との関連表 (KAHOT) を求め、格助詞を含む 1,200 文の実験を行った。その結果、推定した補助語の正解率が約 95% であり、本手法の有効性を確認することができた。

1. はじめに

過去、日本でも、中国でも、英語を主要な対象として機械翻訳の研究開発を行うことが多いが^{1)~10)}、日中両言語間の機械翻訳に関しては本格的な研究が開始されたばかりであり、この分野は日英と中英言語間の機械翻訳などと異なり、未開拓の部分が極めて多い^{11), 12)}。

中国語では文を構成する単語間のあらゆる文法関係は“詞”(独立品詞)の語順と“辞”(補助語)との二つの手段だけで表されている。機械翻訳の規則を作製する際に考慮されなければならないのも、まさにこの点である^{13)~15)}。われわれは、日中両言語の特徴に基づいて、日中機械翻訳システムの研究開発を進めている¹¹⁾。日中機械翻訳を行うためには、上述のように、①独立品詞の語順の確定、②補助語とくに日本語格助詞に対応する補助語の推定が必要と考えられる。筆者らは、コード方式によって、特に常用文型を重視して①の研究を進めており^{16)~18), 23)}、本論文では②について述べる。

一般に、日本語の格助詞に対応する中国語補助語は一对多である¹⁹⁾。例えば、

例文①: 彼は 飛行機で 東京に 行く

訳文: 他乗飛機去東京

例文②: 彼は 食堂で 餃子を 食べる

訳文: 他在食堂吃餃子

の中に格助詞“で”に対応する補助語はそれぞれ“乗”

と“在”である。前者は交通手段を表し、後者は動作場所を表す。さらに、

例文③: 彼は 自転車で 東京に 行く

訳文: 他騎自行車去東京

では、格助詞“で”に対応する補助語は“騎”である。例文①と例文③では、同様に交通手段を表すが、下位区分によって異なる補助語で表されている。これは中国語の特徴の一つである。それゆえ、複数補助語候補から正しい補助語を推定することは日中機械翻訳を行うとき解決しなければならない問題である。

本論文では、意味属性による日本語格助詞に対応する中国語補助語を推定する手法を提案する。本論文において、動詞の意味属性は IPAL²⁰⁾ の意味分類を採用し、名詞の意味属性は文献 21) を参照して設定したものを採用している。われわれは、教科書、文献、文法書など、約 12,000 文から格助詞およびその関連情報を抽出してこの情報を分析し、日本語格助詞と中国語補助語との関連表 (KAHOT と呼び、3 章で述べる) を求めた。これを用い、格助詞を含む 1,200 文について実験を行った結果、補助語の正解率は約 95% であり、本論文で提案する中国語補助語の推定手法の有効性を確認することができた。以下、第 2 章において、格助詞の多義性と多訳性、第 3 章において、補助語の推定アルゴリズム、第 4 章において、この方式の評価と考察について述べる。

2. 格助詞の多義性と多訳性

2.1 格助詞の多義性と多訳性

上に述べたように、日本語の格助詞は、意味上では多義性があり、中国語の訳語(補助語)では多訳性が

† An Algorithm for Estimating Chinese Supplementary Words in Japanese-Chinese Machine Translation System by FUJI REN, YOSHIKAZU MIYANAGA and KOJI TOCHINAI (Faculty of Engineering, Hokkaido University).

†† 北海道大学工学部

表 1 格助詞の補助語候補
Table 1 Chinese supplement words of Japanese case particles.

格助詞	意 味	補 助 語	格助詞	意 味	補 助 語
が	主体	∅	と	対象, 共同者	和
から	事物・空間・時間の始点	从, 自, 由	に	位置, 場所	在, 于
	事物の来源, 材料	由, 用, 以		時間	在, 于, 從
	原因, 理由	因為, 由于		対象	向, 給, 為, 于
	判断の出発点, 根拠	根拠, 从, 按照		事物, 状態の変化と結果	当, 為
で	場所, 位置	在, 于		目標, 目的	為, 為了, 要, 使
	範囲, 範疇	在		去向	到, 往, 向
	方法, 手段, 材料	用, 以		比較の基準	比, 与
	原因, 理由	因為, 由于		計量, 比例の基準	每, 与~成
	交通工具	乘, 坐, 騎		原因, 理由	因為, 由于
	時間, 期限	用, 需要		他	∅
	数量, 価値	花	の	主体	∅
	立場, 身分, 資格	以		修飾	的
へ	方向, 目標	往, 向, 到	まで	時・空・数量の終点と限度	到, 止
	対象	給, 向	を	対象	∅, 把
より	比較の基準	比		[自動詞] 空間・場所・位置の移動, 離開	于, 在, 从
	事物・時間・場所の始点	自, 由, 從			

[∅: 補助語なしを意味する]

ある。なお、本論文では、多様性と多訳性の区別を必要としないので、両方とも多義性と呼ぶ。例えば、以下の例文④、例文⑤、例文⑥の格助詞“で”の場合、示した訳文中の [] 内に示されるように中国語の補助語を決定しなければならない。

例文④: 彼は 病気で 学校を 休む。

訳文: 他 [因] 病 休 学

例文⑤: 彼は 5元で この和服を 買った。

訳文: 他 [花] 5元 買了 這件和服

例文⑥: 彼は 上海で この和服を 買った。

訳文: 他 [在] 上海 買了 這件和服

上例のように、日本語文の格助詞“で”に対応する中国語の補助語は各々“因”、“花”、“在”である。本論文では、これを格助詞の多義性といい、“因”、“花”、“在”などを格助詞に対応する中国語の補助語候補

(略記は補助語候補である) という。また、補助語候補から適当な補助語を選択することを補助語の推定という。

主要な日本語の格助詞に対応する中国語の補助語の候補を整理した結果を表1に示す。表1から補助語候補が多いものは格助詞“で”と“に”であることがわかる。

2.2 格助詞に対応する中国語補助語分布

われわれは、補助語の推定アルゴリズムをつくるために、表2に示される日本語教科書、科学技術文献、辞書などから約12,000文を調べて、格助詞の情報を抽出した。格助詞“で”と“に”の補助語の分布状況は各々表3と表4に示すようになった。なお、抽出の規則として以下の三つの点があげられる。

[1] 格助詞の直前部分は、必ず体言あるいは体言の特性

をもつ文である。用言の場合は解決しているので¹¹⁾、それを除いて体言の場合のみを考える。

例えば、以下の例文⑦の格助詞“と”の情報を抽出するが、例文⑧の“と”の情報を抽出しない。言語学上では例文⑧の“と”を接続助詞ということがある。

例文⑦: 一番よい 時季は 五月と 十月だ

訳文: 最好的時節是五月和十月

例文⑧: 雨が 降ると 道が わるくなる

訳文: 一下雨路就不好走

[2] 格助詞で常用文型(4.1節の脚注を参照)を構成する場合はその格助詞を抽出しない。これはわれわれのシステムでは前もって常用文型の処理を行い、対応する補助語を含めて処理している²³⁾からである。例えば、

体言 に ちがいない (1)

体言 に もかわらず (2)

などは常用文型なので、ここでは抽出を行わない。

[3] 文法関係による格助詞の格関係が変化を生じ

表 2 抽出作業に使用した資料
Table 2 Collected data for the experiment.

No.	教科書あるいは文献名	著 者 名	出 版 社, (年)	ページ数
1	日語 1 (理工科用)	周 炎 輝	高等教育出版社, (1988) 第二版	244
2	日語 2 (理工科用)	周 炎 輝	高等教育出版社, (1988) 第二版	320
3	日語 3 (理工科用)	周 炎 輝	高等教育出版社, (1988) 第二版	231
4	日本語 2 (東大外)	王 二 貴, ほか	山西人民出版社, (1986) 第一版	580
5	日本語 3 (東大外)	王 二 貴, ほか	山西人民出版社, (1986) 第一版	570
6	中日交流標準日本語 2	光 村 等	人民教育出版社, (1990) 第一版	448
7	科学と人類	張 西 祥	機械工業出版社, (1987) 第一版	167
8	科技日語 1	孫 明 久	科学普及出版社, (1984) 第一版	230
9	新日漢辞典	大 連 外	遼寧人民出版社, (1979) 第一版	2678
10	日語語法	王 日 和	商 務 印 書 館, (1981) 第一版	645

表 3 格助詞“で”の補助語の分布
Table 3 The supplement words of the case particle “で.”

No.	補 助 語	頻 度	頻度順
1	在	157	1
2	用	72	2
3	Ø	49	3
4	以	17	4
5	由於	6	5
6	由	6	5
7	坐	4	6
8	因	3	7
9	從…來說	2	8
10	花	2	8
11	驅	1	9
12	從	1	9
13	透過	1	9
14	照	1	9
15	懷着	1	9
16	按照	1	9
17	在…來說	1	9
18	通過	1	9
19	利用	1	9
20	乘	1	9
21	按	1	9
22	騎	1	9
23	出於	1	9

表 4 格助詞“に”の補助語の分布
Table 4 The supplement words of the case particle “に.”

No.	補 助 語	頻 度	順 位
1	Ø	165	1
2	在	142	2
3	到	18	3
4	於	10	4
5	対	7	5
6	与	6	6
7	向	6	6
8	成	5	7
9	毎	4	8
10	給	2	9
11	在…里	2	9
12	在…中	2	9
13	按	2	9
14	和	2	9
15	在…上	2	9
16	對於	1	10
17	為	1	10
18	離	1	10
19	入	1	10
20	地	1	10
21	了	1	10
22	跟	1	10
23	着	1	10
24	比	1	10

る場合に、その格助詞の抽出をしない。例えば、
例文⑨：彼は 先生に 褒められる

訳文：他被老師表揚

のように、“に”に対応する補助語“被”は受動態処理によって自動的に生成されている¹¹⁾。

なお、日本語文法において処理にくい“が”と“は”については、表 1 に示すように、中国語の補助語が“Ø”であり、ほとんど 1 対 1 なので、これについては問題はない (“は”は表 1 に示さないが、“が”と同じく補助語が“Ø”である)。一方、中国語補助語を

推定するときに、推定しにくいものが“で”と“に”である。それで、本論文では、この2種の格助詞を重点として述べるが、本手法は他の格助詞にも完全に適応できると考えられる。以下に、本論文で提案する補助語の推定アルゴリズムを述べる。

3. 補助語の推定アルゴリズム

3.1 格助詞の類別

2.1 節で示した例④, ⑤, ⑥では、補格助詞“で”の直前の体言の意味属性によって補助語を推定できる。例④の場合、“病氣”という体言はある原因の意味属性をもつので、補助語は“因”であり、例⑤の場合、“5元”はお金の数という意味属性をもつので、補助語は“花”である。また、例文⑥の場合、“上海”は場所という意味属性をもつので、補助語は“在”である。同じく、例文①, ③の場合、交通手段の下位区分によって補助語を推定できる。しかしながら、以下の例⑩, ⑪のように体言だけの意味属性では正しい補助語を推定できない場合がある。

例文⑩: 彼は 北京に 住んでいる

訳文: 他住在北京

例文⑪: 彼は きのう 北京に ついたばかりです

訳文: 他昨天才到北京

例文⑩, ⑪中の格助詞“に”の直前の体言は意味属性が同じであるが、後ろの動詞の意味属性が異なるので、補助語が異なる。すなわち、もし後ろの動詞が存在(例えば、ある、いるなど)、生活(例えば、くらし、住むなど)を意味するなら、“に”の補助語は“在”であり、それが移動(例えば、行く、つくなど)を意味するなら、“に”の補助語は“φ”である。

一般に、

体言 i 格助詞 i …用言 j <1>

のような日本語文では、 $\{m(\text{体言 } i), m(\text{用言 } j)\}$ の組で補助語を推定できると考えられる。ここで、 $m(x)$ は x の意味属性を表す。また、一部の格助詞では $m(\text{体言 } i)$ だけでも推定できる。本論文では前者をA類格助詞、後者をB類格助詞という。われわれは2.2 節で抽出した情報から格助詞の分類を行い、その

表 5 格助詞の類別

Table 5 The classification of case particles.

類別	格助詞
A類	に, へ, より, から
B類	が, を, の, で, と, まで

結果を表5に示す。

なお、本システムでは意味属性を通常の用法より広い意味で用いており、体言と用言の意味分類はもちろん、ある特殊品詞の組み合わせ、例えば、「Nの中」、「Nの方面」なども含めている。ここで、Nは体言を意味する。この組合せの意味属性はNの意味属性に関係なく、最後の品詞だけに関係する。このような組合せに対して意味属性を与えることは補助語を推定する際に、非常に効果的であると考えられる。また、本論文の意味属性において、上位、下位の区別は必要としない。例えば、0300 (交通手段) と 0301 (3類交通手段) において、通常は 0301 は 0300 の下位区分であるが、本手法では、同一のレベルで処理する。また、品詞あるいは文節の意味属性は複数の値をもつことがある¹⁹⁾。

3.2 KAHOT の構造

補助語を推定する際に、体言あるいは用言の意味

表 6.1 KAHOT の構造と格助詞“で”の情報

Table 6.1 Organization of the KAHOT (1).

格助詞 (で)	類別 (A, B)	補助語候補数 (15)
順番	M (体言 i)	M (用言 j) 補助語コード記号
1	場所 0100	— 在 T001
2	交通手段(3類) 0301	— 騎 T002
3	交通手段(2類) 0302	— 乗 T003
4	交通手段(1類) 0303	— 坐 T004
5	原因 0700	— 因 T005
6	お金 0801	— 花 T006
⋮		

表 6.2 KAHOT の構造と格助詞“に”の情報

Table 6.2 Organization of the KAHOT (2).

格助詞 (に)	類別 (A, B)	補助語候補数 (17)
順番	M (体言 i)	M (用言 j) 補助語コード記号
1	場所 0100 地点 0101	存在の動詞 V010 静止の動詞 V011 数量の形容詞 V030 在 N002
2	地点 0101	移動性 V003 φ N001
3	地点 0101	集積性 V003 在 N002
4	—	趨向性 V004 向 N003
5	周期性 0600	分配 V041 毎 N004
6	—	比較 V043 与 N005
⋮		

属性などの本手法に必要な情報は格助詞補助語関連表 (KAHOT と呼ぶ) に登録される。KAHOT の構造を、格助詞 “で” については表 6.1, 格助詞 “に” については表 6.2 にそれぞれ示した。

3.3 アルゴリズムの流れ

本論文で提案するアルゴリズムの流れを図 1 に示す。

図 1 では、②番の出口は補助語が推定できたことを表し、①番の出口は推定できなかったことを表す。①は品詞の意味属性を決定する情報が不足である場合と、KAHOT の情報が足りない原因である。実験システムでは、①番の出口の場合に、各々の格助詞の補助語分布から、出現頻度の一番高い補助語を取る。例えば、“で” の場合に補助語 “在”, “に” の場合に補助語 “φ” を決定する。

以下に、例文を用いて推定手順の概要について述べる。なお、各変数の意味を次に示す。

KAWD: 格助詞名
 NWORD: 格助詞の直前の単語
 IMI(I): 単語の意味属性, $I=1\sim N$
 N: 単語の意味属性値の数
 IMIN: 単語の意味属性
 IMIK(J): KAHOT に登録された J 番目の意味属性

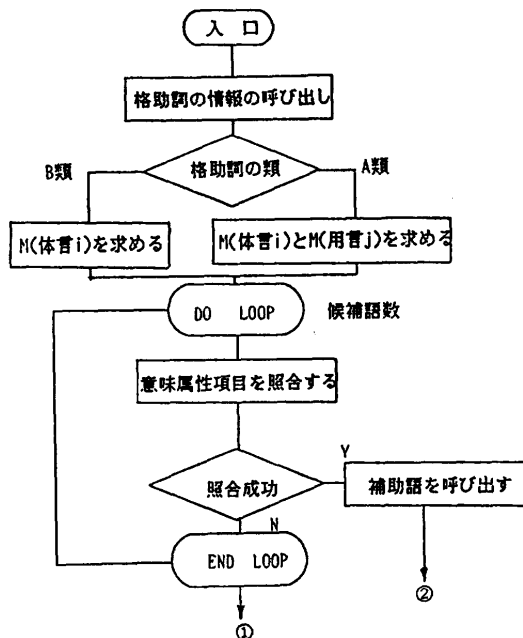


図 1 格助詞の補助語の推定の流れ
 Fig. 1 Flowchart of the Algorithm.

CODE(J): 補助語のコード
 NCODE: 推定した補助語のコード
 SUWD: 推定した補助語
 SEIMI: サブルーチン名
 その機能は、ある単語 “NWORD” の意味属性を求め、配列 “IMI” に置くことである

例文⑩: 鈴木さんは 明日 船で 行く

ここで、

格助詞名 KWORD=“で”

格助詞直前の単語 NWORD=“船”

STEP 1 サブルーチン SEIMI を呼び出し、IMI と N を求める。

$N=2$

IMI (1)=0303 (交通手段, 1 類)

IMI (2)=0313 (交通手段, 水面だけに使う)

STEP 2 LOOP 1 を実行する。これは単語の複数の意味属性に対し一つずつ処理する。

IMI の第一番目の値を IMIN に置く

IMIN=IMI(1)=0303

STEP 3 LOOP 2 を実行する。これは IMIN を KAHOT 中の項目と照合を行う。

表 6.1 に示した KAHOT では、 $J=1\sim 3$ では照合が失敗した。

$J=4$ の時に

IMIK (4)=0303=IMIN

の条件を満足するので、照合が成功した

STEP 4 推定した補助語コードを求める

NCODE=CODE (4)=“T004”

STEP 5 推定した補助語

SUWD=SUWDS (T004)=“坐”

最後の訳文: 鈴木明天坐船去

例文⑪: 鈴木さんは 明日 自転車で 行く

この例で、

格助詞名 KWORD=“で”

格助詞直前の単語 NWORD=“自転車”

“自転車” の意味属性は “0301” (交通手段, 3 類) なので、KAHOT の第 2 番目の項目との照合が成功し、“騎” という補助語が推定される。

最後の訳文: 鈴木明天騎自転車去

A 類格助詞の補助語の推定手順は B 類と同じであるが、照合を行っている条件は次の二つである。

条件1: 格助詞の直前の体言の意味属性はKAHOT中の項目M (体言 i) と一致する

条件2: 格助詞と関連する用言の意味属性はKAHOT中の項目M (用言 j) と一致する

3.1 節で示した例文⑩で、

格助詞“に”の直前の体言=“北京”

格助詞“に”と関連する用言=“住んでいる”

であり、各々表 6.2 の第一項目の条件を満足するので、照合が成功し、補助語“在”が推定される。

例文⑩の場合に、

格助詞“に”の直前の体言=“北京”

格助詞“に”と関連する用言=“つく”

であり、表 6.2 の第一項目の第1条件M (体言 i) を満足するが、第2条件M (用言 j) を満足しないので、一番目との照合が失敗する。つづいて、第二項目と照合する。ここで、二つの条件を同時に満足するので、照合は成功し、補助語“〇”が推定される。

4. 評価と考察

4.1 日中機械翻訳について

本論文で前提としている機械翻訳方式は日本語と中国語の構造の対照分析に基づく構造変換である。構造変換法自体は非常に実用的で有効な手法であると考えられるが²²⁾、単にこの方法を用いただけでは膠着語である日本語と孤立語である中国語の間の固有の表現のギャップを解決できない場合が多い。中国では日本語の固有の表現、例えば、“まるで…ようだ”、“けっして…ない”、“しか…ない”などを常用文型^{*}と呼んでいる。一般に中国では、日本語の学習に際してまず日本語のいくつかの格助詞を学び、その後、常用文型を中心に学習する。とくに日本語文を翻訳する際に、常用文型については詳細な文法解析を行わず、直接に中国語訳文を得ている。また、このようにして得られた中国語訳文は常用文型に対応して選択される補助語によりアスペクトとモダリティも満足させる。われわれはこの中国人の日本語学習の方法が日中機械翻訳に効果的であると考え、日本語文を基本文と常用文型に分解し、基本文に対して構造変換を行い、常用文型に対してあらかじめ用意した訳文関数^{**}で解析する手法

* 中国語ではこのような表現を“慣用句型”(guanyong-juxing)と呼んでいる。しかし、ここでの“慣用句型”は日本語における“慣用句”とは完全には一致しないので、本論文では“慣用句型”の日本語訳として“常用文型”という語句を用いる²³⁾。

** 訳文関数は日本語の固有の表現と中国語の固有の表現の対応関係である。この対応関係は訳語と語順情報を含む。例えば、常用文

を用いて、日中機械翻訳システムの構築を行っている²⁴⁾。また、常用文型と関連する要素の意味属性を利用し、常用文型の多義性と多訳性を解消する手法を提案し¹⁷⁾、実験によりその有効性を確認した。上記翻訳手法と多義性の解消方式の記述は別の機会に譲る。次に、本論文における提案した中国語補助語の推定手法の有効性を確認する実験について述べる。

4.2 評価実験

表2に示した資料から無作為に抽出した格助詞を含む1,200文に対する実験を行い、本アルゴリズムの妥当性を評価した。ここで、実験は翻訳システム全体と分離して行い、評価データは補助語のみを指標として行った。結果の客観性を調べるため、中国人留学生3人に無作為に抽出した600文の結果を評価させた。それぞれの評価結果を表7に示す。

4.3 考察

表7において、Iは教科書、文献の中に現れる補助語と本手法で推定した補助語と比較して得られた結果であり、IIは言語専攻の中国人留学生に評価された結果であり、IIIは工学専攻の中国人留学生二人で評価された結果であり、IVは著者の一人によって評価された結果である。表7から、Iの正解率が一番低く、IIIとIVは大体同じであることがわかった。また、IIはIII、IVよりややきびしい評価であった。Iの結果と他の結

表7 実験評価結果
Table 7 Results of the experiment.

Tester	評価結果	
	正解数	正解率 (%)
I Texts	497	82.8
II Chinese student C	547	91.1
III Chinese students A, B	571	95.1
IV Author	568	94.7

型“…けっして…ない”の訳文関数は次の式<F>である。[主]は主語要素を意味する。

けっして…ない→[主]绝不… <F>

*** 例文で翻訳の手順を示す。例文の中で“けっして…ない”は常用文型である。

例文: 真由美はけっしてレストランで餃子を食べない。

①分割 基本文: 真由美はレストランで餃子を食べる。

常用文型: けっして…ない

②基本文の構造変換

{真由美, レストラン, 餃子, 食べる} (日本語構造) →

{真由美, レストラン, 食べる, 餃子} (中国語構造)

(真由美, 西餐廳, 吃, 餃子)

③補助語の推定: は→“〇”, で→“在”, を→“〇”

④常用文型の訳文関数: [主] 绝不…

⑤訳文合成: 真由美绝不_〇在西餐廳吃餃子。

果との大差の原因としては以下の2点が考えられる。

a. 各々作者によって異なる補助語を採用し、一意な訳語を選択しない。例えば、表3中に補助語“按”、“照”、“按照”としては、ほとんど区別がないが、作者の好みによって選択しており、本システムでは“按照”を登録している。

b. 同じ文でも異なる補助語を採用できる。例えば、“私たちはそれを最初の日々のような、慕わしい気持ちで、肩を押しつけ合ったまま、たたずんでいた”(表2. No. 5, p. 280, 原著: 風立ちぬ, 堀辰雄作品から)、この格助詞“で”に対応する補助語としては、作者は“懷着”を採用したが、本実験システムでは、“以”をその補助語として KAHOT に登録している。

それで、本論文では、IVを評価の基準としている。

われわれが実験した1,200文の結果をまとめると、補助語を正しく推定できた正解率が約95%であり、その中で、A類の正解率が約92%、B類の正解率が約97%であり、本論文で提案した補助語の推定手法の有効性が確認された。なお、不正解の場合には、主に次の3種がある。

I. 意味属性の細分類の不足

実験における意味属性は文献20, 21)を参照して設定したものであるが、KAHOTにおける情報はほとんど何の方法論も用いず資料から適当に翻訳例を集めることによって作成したものである。また、最初の実験では、意味属性の70種類を用いて実験を行った。その結果は正解率がわずかに74%であった。その後、不正解の文を分析し、意味属性の数を増し、KAHOTの情報を修正して繰り返し実験を行った。最後の結果は意味属性を137種類登録した時点で得られた結果である。このように、本手法の正解率は体言と用言の意味属性の与え方の細かさに依存する。

次に不正解の1例を示す。例文⑭Bでは格助詞“に”に対応する補助語が正しく推定できなかった。

例文⑭A: 張先生が私に日本語を教える。

例文⑭B: 私は張先生に日本語を習う。

実験システムでは、“教える”と“習う”の意味属性として次の3種類が登録されている。

- (1) 他動詞
- (2) 動作性
- (3) 授受性

また、格助詞“に”に対して KAHOT に以下の情報が登録されている。

M (体言 i): 人類	〈条件1〉
M (用言 j): 授受性	〈条件2〉
中国語補助語: $\emptyset$	〈結果〉

すなわち、例文⑭A, Bでは、“に”の直前体言と後属用言の意味属性が上記の条件〈1〉, 〈2〉をそれぞれに満足するので、補助語“ $\emptyset$ ”が推定される。この結果は例文⑭Aでは正しいが、例文⑭Bでは誤りであり、後者の正解は“跟”である。このような誤りは意味属性の細分により解決できると考えている。例えば、上例の場合に、意味属性“授受性”を“内向授受性”と“外向授受性”に細分すれば、A, Bともその補助語を正しく推定できる。むしろ、相応する KAHOT の情報を追加しなければならない。

II. 体言の意味属性は通常以外に使用すること

本論文において、推定の効率のために、一部の資料を調査してから、格助詞“で”をB類とした。B類格助詞では、“で”の直前体言の意味属性だけを用いて補助語を推定しているが、つぎのような誤りがあった。

例文⑮: 王鉄牛は自転車で凶悪な虎をなぐる。

通常に、“自転車”の意味属性は“交通手段”であるが、ここで、“自転車”の意味属性は“交通手段”ではなく、“道具、武器”として使われる。それで、推定した補助語“騎”は間違えた。正しい補助語は“用”である。このようなことは表2の資料を調べて数が少ないが、大量化した場合に実データの収集および処理手法の改良が必要であると考えられる。

III. 登録の情報の不足で推定できないこと

これは設定した体言または用言の意味属性と登録した KAHOT の情報の照合が失敗したことである。

以上により、広範な分野にわたる大量化した場合に本実験における意味属性の数と KAHOT の情報はまだ不十分であることがわかった。より精密で標準化された KAHOT を確立することが必要であると考えられる。

また、本論文では、格助詞のパターンを〈2〉式として考えた。

体言 i 格助詞 i … 用言 j 〈2〉

すなわち、

{T_i, K_i, Y_j} 〈3〉

の構造を用い、抽出、照合、推定を行っているが、実際には、

体言 i₁ 格助詞 i₁ 体言 i₂ 格助詞 i₂ …

体言 i_n 格助詞 i_n 用言 j 〈4〉

のようなパターンもある。すなわち、

$$\{T_{i_1}, K_{i_1}, T_{i_2}, K_{i_2}, \dots, T_{i_n}, K_{i_n}, Y\} \quad \langle 5 \rangle$$

という構造が存在している。ここで、Tは体言を意味し、Kは格助詞を意味し、Yは用言を意味する。以下に示す例文⑯、⑰はこのような例である。

例文⑯：彼女は何から何まで母親にそっくりだ。

例文⑰：団長は三時から五時まで会議室にいる。

この構造に対して次の式(6)のように分解してそれぞれの格助詞について相応する補助語を独立に推定する手法が考えられる。

$$\{T_{i_k}, K_{i_k}, Y\} \quad k=1, \dots, n \quad \langle 6 \rangle$$

これにより、B類格助詞の補助語の推定は問題がほとんどないが、A類格助詞に対してその実用性についてさらに検討する必要があると考えられる。

5. おわりに

本論文では、日中機械翻訳システムにおける日本語格助詞の中国語補助語の推定手法を提案した。これは、大量の教科書と文献から格助詞と関連の情報を抽出し、格助詞をA類とB類に分けて、品詞の意味属性を用いて正しい補助語を推定するものである。また、格助詞を含む1,200文の実験結果から、推定した補助語の正解率が約95%であり、本手法の有効性を確認することができた。今後、より精密で標準化されたKAHOTを確立し、翻訳システムに組み込む処理などを予定している。

謝辞 日ごろ有益なご討論、ご助言をいただく研究室各位に感謝いたします。また、実験システムの構築を進めるにあたり種々ご助言をいただいた斎川勝男技官に感謝の意を表します。

参 考 文 献

- 1) 長尾 真：機械翻訳サミット，オーム社 (1989)。
- 2) 野村浩郷，田中穂積：機械翻訳，bit 別冊，共立出版社 (1988. 9)。
- 3) 高松 忍，西田富士夫：動詞パターンと格構造に基づく英日機械翻訳，信学論(D)，Vol. J 64-D，No. 9，pp. 815-822 (1981)。
- 4) 石崎 俊，内田裕士：多言語間翻訳のための中間言語について，情報処理学会研究報告，NL-70-3 (1989)。
- 5) 鈴木康広，柄内香次：語の接続関係を利用した機械翻訳システム，情報処理学会論文誌，Vol. 29，No. 4，pp. 342-349 (1988)。
- 6) 辻井潤一：機械翻訳システム，スペクトラム，Vol. 1，No. 11，pp. 48-57 (1988)。
- 7) 劉 涌泉：机器翻譯淺說，中国語文 1958，12月

号，pp. 575-577 (1958)。

- 8) Liu, Z.: An Introduction to JFY-11 English-Chinese Machine Translation Algorithm, ZHONGGUO YUWEN (中国語文), May, pp. 216-220, July, pp. 279-285 (1981)。
- 9) Chen, Z. X. and Gao, Q. S.: Intelligent English-Chinese Machine Translation System, Science in China (Series A), Vol. 32, No. 8, pp. 25-32 (1989)。
- 10) Wang, K., Guo, W. and Zhang, L.: HH-87 Machine Translation System, Vol. 12, No. 6, pp. 472-476 (1989)。
- 11) 任 福継，宮永喜一，柄内香次：コード方式日中機械翻訳の実験システム JCMTS の概要，情報処理学会研究報告，Vol. NL 72-7，No. 42，pp. 1-8 (1989)。
- 12) 李 統漢：“日譯中”的中文表達，日語學習与研究，中国北京對外經濟貿易大学《日語學習与研究》編輯委員會(編)，pp. 31-35，中国国际圖書貿易总公司 (1989. 4)。
- 13) 望月八十吉：中国語と日本語，光生館 (1974)。
- 14) 今富正巳：中国語⇔日本語翻訳の要領，光生館 (1976)。
- 15) ソフローノフ M. B. (橋本萬太郎(訳))：中国語機械翻訳の一般原理，中国語学，pp. 13-20 (1961)。
- 16) 任 福継，宮永喜一，柄内香次：日中機械翻訳システムにおける慣用表現のエンコードおよびデコード，信学技報，NLC 89-30，AI 89-56，pp. 39-46 (1989)。
- 17) 任 福継，宮永喜一，柄内香次：慣用表現の訳文生成について，信学大全，D 92 (1990. 3)。
- 18) 任 福継，范 莉馨，宮永喜一，柄内香次：JCMTS における慣用表現の解析手法，信学技報，NLC 90-1，pp. 9-16 (1990. 5)。
- 19) 任 福継，宮永喜一，柄内香次：格関係と意味属性による補助語の推定手法，北海道支部連大論文集，pp. 333-334 (1990. 10)。
- 20) IPAL，計算機用日本語基本動詞辞書，情報処理振興事業協会技術センタ (1987)。
- 21) 荻野孝野：日本語の意味分類体系，計量国語学，Vol. 16，No. 3，pp. 95-112 (1987)。
- 22) AI 総覧，(第7章機械翻訳)，フジ・テクノシステム/エス・ディ・シー (1987)。
- 23) 任 福継，宮永喜一，柄内香次：日中常用文型機械翻訳システム，信学論(D-II)，Vol. J 74-D-II，No. 8，pp. 1060-1069 (1991)。

(平成2年11月29日受付)

(平成3年9月12日採録)

任 福維 (正会員)

1982年中国北京郵電学院電信工
程部計算機と通信専攻卒業。1985年
同大学院計算機応用専攻修士課程修
了。1986年中国科学院博士課程入
学。1987年中退来日。1991年北海道
大学工学研究科電子工学専攻博士課程修了。工学博士。
計算機科学，自然言語処理，特に機械翻訳の研究に従
事。

栃内 善次 (正会員)

昭和14年生。昭和37年北海道大
学工学部電気工学科卒業。昭和39
年同大学院工学研究科修士課程修
了。現在同工学部電子工学科教授。
工学博士。自然言語処理，音声情報
処理および信号処理プロセッサなどの研究に従事。電
子情報通信学会，日本音響学会各会員。

宮永 喜一 (正会員)

1956年生。1981年北海道大学工
学部電子工学専攻修士修了。工学博
士。現在，北海道大学工学部電子助
教授。並列計算機システム，ディジ
タル信号処理等の研究に従事。電子
情報通信学会，日本音響学会，IEEE 各会員。