

概要

高品質な翻訳の実現を目指して、大規模な文型パターン辞書を用いた翻訳方式の研究が行われている [1]。日英翻訳では、すでに大規模の網羅的な結合価パターン辞書 [2] が開発され、機械翻訳において、結合価パターン方式が大変有効であることが報告されている。しかし、結合価パターン方式は日英翻訳の研究が多く、その他の言語族についての検討がまだない。

そこで、本研究では、日中翻訳を対象に想定して、日中翻訳における結合価パターン方式の有効性を評価するため、結合価パターン辞書を部分的に作成し、作成した結合価パターンを用いて、翻訳実験を行う。また、翻訳実験の結果により、結合価パターン翻訳方式の効果を評価する。

具体的には、結合価パターンによる方式は、語義数の多い用言の訳し分けにおいて特に効果が期待される方式である。通常、語義数の多い用言は、使用頻度も高いと考えられるので、本研究では、使用頻度の高い日本語結合価パターンを対象に日中結合価パターン辞書を作成し、さらに、日中結合価パターンを用いて翻訳能力を検討する。

翻訳試験の結果により、翻訳文の全体には、評価値 A が 80(160/200) %，評価値 B が 7.5(15/200)% を得られた。翻訳実験を通して、日中翻訳における結合価パターン翻訳方式は大変有効な方法が分かった。

検討の結果では、作成した中国語結合価パターンは、日英翻訳のための日本語結合価パターンを利用した。そのために、一つの日本語結合価パターンに対して、複数の中国語結合価パターンが対応する場合がある。この原因で、一つの中国語結合価パターンに定義出来ず、翻訳全体の精度は、約 15%(30/200) 悪くなった。問題を解消するために、今後の課題として、日本語語彙大系の「意味属性体系」を細かく、及び、追加する必要性があると考えている。

また、日中結合価パターン辞書を完成して、機械翻訳に応用すれば、日中機械翻訳の全体も、高い翻訳品質を得られると考えている。

目次

1	はじめに	2
2	日中結合価パターンの部分試作	3
2.1	日本語語彙大系の結合価パターン	3
2.2	中国語結合価パターンの作成方法	5
2.3	中国語結合価パターンの作成結果	6
3	結合価パターンによる翻訳能力の評価	7
3.1	結合価パターンを用いた翻訳方法	8
3.2	試験用入力文	10
3.3	評価基準	11
3.4	評価の例	12
3.5	評価結果	16
4	考察	17
4.1	日本語結合価パターンのカバー範囲の問題	17
4.2	結合価パターン方式の限界を超える問題	23
5	おわりに	24

目次

1	日中翻訳実験の流れ	8
2	選択された日本語単文例	10
3	「明日」の語順	23

表目次

1	日中翻訳実験の結果	16
2	「パターン：でる」の意味属性調査表	17
3	「パターン：でる」に対して作成できる中国語結合価パターン	18
4	「パターン：ひどい」の意味属性調査表	19
5	「パターン：ひどい」に対して作成できる中国語結合価パターン	19
6	「パターン：綺麗」に対して作成できる中国語結合価パターン	21

1 はじめに

機械翻訳システムは，ここ四十年ほどの間に，世界各地で，大規模なプロジェクトが多く行われる．また，機械翻訳システムには，パターン翻訳方式，統計翻訳方式と用例翻訳などが多く使用されている．従来の研究によって，これらの翻訳手段は効果があることが分かった．しかし，訳文はまだ理想の品質になってない．

ここ数年，言語表現の意味を等価的に変換する方式として，パターンを用いた翻訳方式に着目した研究が行われている [1]．しかし，これまで，日英翻訳の研究が多く，その他の言語族についての検討が少ない．特に，日英機械翻訳では，効果がある事を証明した結合価パターン翻訳方式についての検討がまだない．そこで，本研究では，日中翻訳を対象として想定して，中国語結合価パターンを試作し，さらに，日中結合価パターンを用いて翻訳能力を実験的に評価する．

2 日中結合価パターンの部分試作

本研究では、すでに開発されている日英結合価パターン辞書の日本語パターンを対象に、対応する中国語結合価パターンを試作し使用する。

ところで、結合価パターンは、体言と用言の意味的な関係をパターン形式で表現したものであり、機械翻訳では、原言語のパターンと対応する目的言語のパターンを対にした結合価パターン対辞書が使用される。パターン対辞書の設計では、原言語のパターンに適合した入力文に対して、目的言語のパターンが一意に決定できるようにする必要がある。このため、参照する日英パターン辞書では、日本語パターンは、適合したすべての日本文が対応する英語パターンを用いて翻訳できるように、カバー範囲が決められている。従って、日中翻訳の場合は、この辞書に登録された日本語パターンは、必ずしも中国語のパターンに1対1に対応しない可能性がある。

そこで、本研究では、作成した日中結合価パターン辞書を用いた翻訳実験から、日中翻訳における日本語パターンの構成法上の問題点とそれを用いた結合価パターン方式の可能性を明らかにする。

2.1 日本語語彙大系の結合価パターン

本論文では、岩波の日本語語彙大系 [2] の結合価パターンを日本語結合価パターン辞書として、用いる。本研究では、使用される日本語結合価パターン辞書を以下に詳しく説明する。

まず、岩波の日本語語彙大系の構成を説明する。岩波の日本語語彙大系は、「意味属性体系」、「単語体系」と「構文体系」の3体系から構成される。

「意味体系」は、日本語の一般名詞、固有名詞、用言の意味的用法を3,000の意味属性で体系化した「意味属性体系」と、意味属性ごとにその属性をもつ単語を集めた「意味属性別単語表」からなる。

「単語体系」は、日本語の意味的用法を3,000の意味属性を用いて定義したもの。一般名詞のほかに使用頻度の高い固有名詞や新聞記事に現れる時事用語なども含む30万語を収録する。

「構文体系」は、日本語の用言 6,000 語の文型を結合価パターンにまとめたもの。一般の文型と慣用的表現の文型を合わせて 14,000 の日本語文型パターンにまとめ、同時にそれに対応する英語の文型パターンを収録する。

これら 3 つの体系を合わせて、日本語結合価パターンとして、機械翻訳、構文解析などよく使用されます。以下は、結合価パターンの記述例である。

なお、[] 内の記述はパターンが可能な文型状態である、() 内の数値は意味属性である

- 例 1：日本語結合価パターン：

N1 "が" N8 "に" 無い [受身不可]

N1 の意味属性：(*全ての意味属性)

N2 の意味属性：(388 場所 533 具体物 1000 抽象)

- 例 2：日本語結合価パターン：

N1 "が" N8 "を" 味わう [動作]

N1 の意味属性：(4 人)

N2 の意味属性：(838 食料 1253 感情)

2.2 中国語結合価パターンの作成方法

本研究では，日中機械翻訳における結合価パターン方式の効果を評価したい．しかし，実験用中国語結合価パターン辞書が存在しないので，予め，人手で作成する．作成する200個の結合価パターンは中国語結合価パターン辞書として，使用する．作成方法を以下のように示す．

まず，日中対訳文を参考し，日本語結合価パターンに対応する中国語結合パターンを作成する．次に，作成した日中結合価パターン対を使用して，参考した日中対訳文を用いて，クローズテストを行う．クローズテストの上に，さらに，対応する中国語結合パターンを補正する．具体的な作成手順を，以下に示す．

1. 日本語結合価パターンの使用頻度調査

日英対訳コーパス(100万件)の中から，日本語単文27万件[3](以下「単文集」とする)を抽出する．単文集をパターンパーサ[6]にかけて，日本語語彙大系[2]の結合価パターン辞書と照合し，使用されている結合価パターンを調べる．さらに，使用頻度を統計する．

2. 対象とする日本語結合価パターンの選択

使用頻度が高い日本語結合価パターン(上位200件)を選択する．

3. 中国語訳語の付与

上記で選択された各日本語結合価パターンに対して，適合する日本文を任意で1文選択し，対応する中国語訳文を付与する．

4. 中国語結合価パターンを作成

上記3.で得られた日中対訳文を参考に，2.で選択した日本語結合価パターンに対応する中国語結合価パターンを作成する．

5. 作成したパターンにクローズテストする

作成した中国語結合価パターンに対して，3.で使用した日中対訳文を入力文として，クローズテストを行う．

6. 補正

クローズテストの結果により，中国語結合価パターンを補正する．

2.3 中国語結合価パターンの作成結果

使用頻度が高い日本語結合価パターンに対して、200 個の中国語結合価パターンを作成した。例を以下に示す。

例 1 :

日本語結合価パターン :

N1 "が" N8 "に" 無い

N1 の意味属性 : (*全ての意味属性)

N2 の意味属性 : (388 場所 533 具体物 1000 抽象)

参考した日本語文 :

黒板 が 教室 に 無い .

付与した中国語の訳文 :

在 教室 没有 黑板 .

作成した中国語結合価パターン :

"在" N8 "里" 没有 N1

例 2 :

日本語結合価パターン :

N1 "が" ひどい

N1 の意味属性 : (3 主体 2055 出来事 1560 行為 2422 抽象的な関係)

参考した日本語文 :

肩こりがひどい .

付与した中国語の訳文 :

肩酸严重 .

作成した中国語結合価パターン :

N1 严重

3 結合価パターンによる翻訳能力の評価

本章では，作成した中国語結合価パターンに対して，二つのことを注目する．

一つ目は，結合価パターンによる方式は，語義数の多い用言に対して，訳し分けの能力を特に期待される方式である．本章では，語義数の多い用言の訳し分けの効果を注目する．

また，既に述べたように，通常，語義数の多い用言は，使用頻度も高いと考えられる．そこで，使用頻度が高い結合価パターンを対象として，翻訳効果を評価すれば，語義数の多い用言に対する訳し分けの能力も分ると考える．

二つ目は，作成した中国語結合価パターンは，日英翻訳のための日本語結合価パターンを利用している．日本語結合価パターンでは，適合した全ての日本語文が対応する英語パターンを用いて翻訳できるように，カバー範囲が決められている．しかし，日本語結合価パターンは中国語結合価パターンと 1 対 1 に対応しない可能である．そのため，入力文に対して，対応する中国語パターンを 1 つに絞り込めない場合があると考えられ，品質の高い中国語が得られるとは限らない．そこで，日中翻訳における日本語結合価パターンの構成法上の問題点を注目する．

上記のことを注目し，日中翻訳実験を試みた．

3.1 結合価パターンを用いた翻訳方法

本研究では，日中翻訳実験の流れを図 1 に示す．

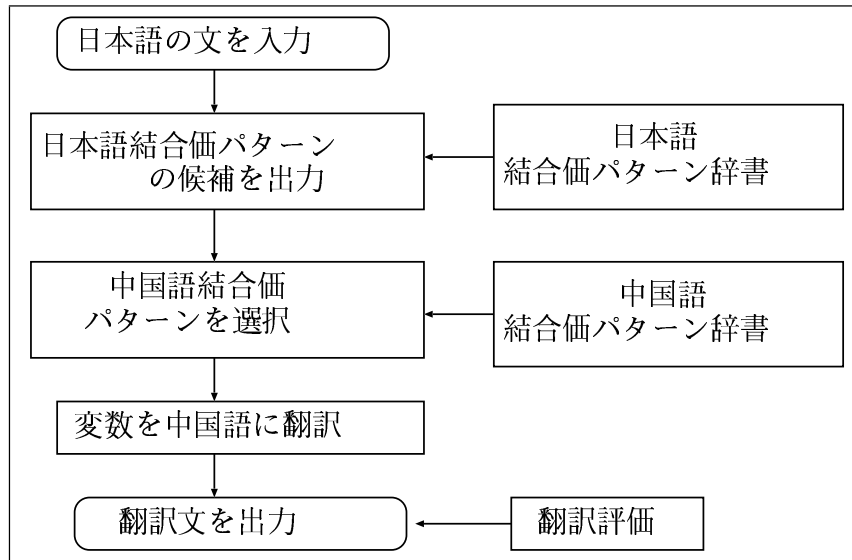


図 1 日中翻訳実験の流れ

具体的に，日中翻訳実験の流れを以下に説明する．

1. 試験用入力を選択：

入力文は日本語単文集 [3] から 200 文を選択する．詳しい選択条件は 3.2 節に説明する．

2. 抽出と適合条件の判断：

入力した日本語文と日本語結合価パターン辞書から当てはまる日本語結合価パターンを取り出す．さらに，入力文が日本語結合価パターン内の変数の意味的な制約条件を満足すれば，両者は適合したと判断される．

3. 中国語結合価パターンの抽出：

選択された日本語結合価パターンに対応する中国語結合価パターンを作成した中国語結合価パターン辞書から取り出す．

4. 変数の中国語の付与：

適合した変数の値（日本語単語）に対して，日中辞典から，パターンで指定された意味属性に適合する訳語（中国語単語）を検索する．日中辞典に適切な訳語が存在すれば，検索は成功する．探索で成功した中国語単語を変数に代入する．上記で得られた訳語を中国語パターンの該当する変数に代入し，中国語訳文を生成する．

5. 訳文の出力：

上記で得られた変数の訳語を中国語パターンの該当する変数に代入し，中国語訳文を生成する．

6. 訳能力の評価：

訳文に対して，「3.3 節の評価基準」を用いて，訳能力を評価する．

次の節から，日中訳実験では，使用された実験用各要素を説明する．

（日中結合価パターン辞書の説明を第 1 章に示す．）

3.2 試験用入力文

本研究では、日本語の単文を利用する。単文の構文構造は単純なので、新しい翻訳方式に対して、テストや検討などが簡単である。約 27 万の日本語単文が収録された単文集 [4] から、以下の条件を満たす日本語単文をランダムに 200 文選択する。

1. 2.3 節で作成した結合価パターンに適合する。
2. 2.2 節において使用した単文ではない。
3. 結合価パターンに含まれるすべての変数が単文に対応する。
4. 日本語単文が用言の終止形で終る。
5. 形容詞句と名詞句は対象内、副詞句を含まない。

選択した日本語の単文例を図 2 に示す。

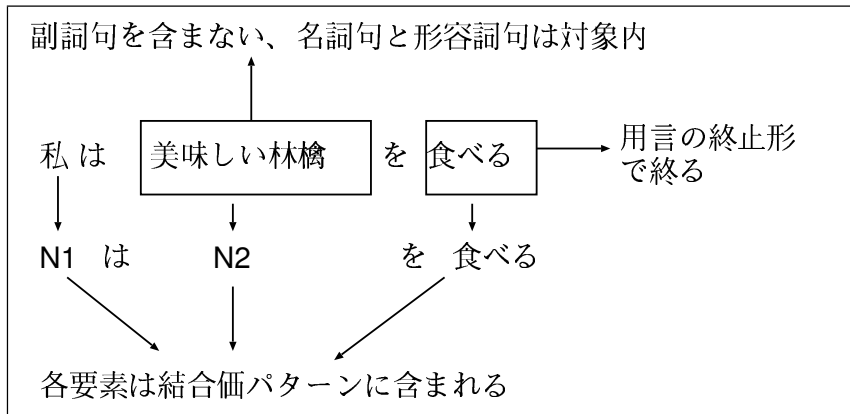


図 2 選択された日本語単文例

3.3 評価基準

本実験では、翻訳文の評価基準は以下の4段階とする。

- A) 文法が正しく、意味が理解できる。
- B) 文法に不自然なところがあるが、意味が理解できる。
- C) 愛味なところがあるが、意味が大体理解できる。
- D) 全く意味が理解できない。

本節の評価基準を用いて、日中翻訳実験を評価する。
(具体的な翻訳例を3.4節に参考してください。)

3.4 評価の例

評価値 A, B, C, D の例を本節の各項に示す.

3.4.1 A 評価の例

例 1:

- テスト文 :
私の家は駅から近い.
- 使用された結合価パターン対 :
日本語パターン : ("近い")
N1 "が" N3 "に" "と" "から" 近い
中国語パターン : ("近")
N1 "和" N3 近
N1 の意味属性 : (*すべての意味属性) N3 の意味属性 : (*すべての意味属性)
- 変数の翻訳 : 家 (*)→家 , 駅 (*)→車站
- 訳文出力 :
我的家和車站近.

例 2:

- テスト文 :
その手法は古い .
- 使用された結合価パターン対 :
日本語パターン : ("古い")
N1 "が" 古い
中国語パターン : ("古老")
N1 古老
N1 の意味属性 : (389 施設 458 地域 706 無生物 1000 抽象)
- 変数の翻訳 : 手法 (1000)→手法
- 訳文出力 :
那个手法古老 .

3.4.2 B 評価の例

例 1:

- テスト文：
三杯の水を飲む。
- 使用された日中結合価パターン対：
日本語パターン：("飲む")
N1 "が" N2 "を"飲む
中国語パターン：("喝")
喝：N1 喝 N2
N1 の意味属性：(4 人 535 動物)
N2 の意味属性：(746 液体 857 飲物)
- 変数の翻訳：水 (746)→水
- 訳文出力：
喝 3 杯的水。
- B 評価の原因：
助詞がないので，文の全体は不自然である。普段は「喝了 3 杯的水。」と言う。

例 2:

- テスト文：
人の心は様々だ
- 使用された日中結合価パターン対：
日本語パターン：("様々だ")
N1 "が" 様々だ
中国語パターン：("各種各樣的")
N1 是 各種各樣的 N1 の意味属性：(* 全ての意味属性)
- 変数の翻訳：心 (*)→心
- 訳文出力：
人的心是各種各樣的。
- B 評価の原因：
文の意味が理解できるが，全体的に不自然である。普段は「人心多變」と言う。

3.4.3 C 評価の例

例 1:

- テスト文：
彼の性格が作品に出る.
- 使用された日中結合価パターン対：
日本語パターン：("でる")
N1"が" N8"に" でる
中国語パターン：("出現")
N1 出現 "在" N8
N1 の意味属性：
(* -2671 暦日以外のすべての意味属性)
N8 の意味属性：(*すべての意味属性)
- 変数値の翻訳：性格 (*) → 性格
- 訳文出力：他的性格出現在作品.
- C 評価の原因：意味を理解しにくい.
普通は「他的性格体現在作品.」と言う.

例 2:

- テスト文：
朝が早い.
- 使用された日中結合価パターン対：
日本語パターン：("早い")
N1"が" 早い
中国語パターン：("早")
N1 早
N1 の意味属性：
(2670 時間)
- 変数値の翻訳：朝 (*) → 早上
- 訳文出力：早上早.
- C 評価の原因：意味を理解しにくい.
普通は「太早」と言う.

3.4.4 D 評価の例

例 1:

- テスト文：
その子供の親はひどい.
- 使用された日中結合価パターン対：
日本語パターン：("ひどい")
N1 "が" ひどい
中国語パターン：("出現")
N1 厉害
N1 の意味属性：(3 主体 2055 出来事 1560 行為 2422 抽象的關係)
- 変数値の翻訳：N1(3):親 → 家长
- 訳文出力：
那个孩子的家长厉害.
- D 評価の原因：意味を理解できない。
普通は「那个孩子的家长無情。」と言う。

例 2:

- テスト文：
彼は嘘がうまい.
- 使用された日中結合価パターン対：
日本語パターン：("うまい")
N1 "は" N2 "が" うまい
中国語パターン：("好")
N1 的 N2 好
N1 の意味属性：(3 主体 535 動物) N2 の意味属性：(760 人工物 1000 抽象)
- 変数値の翻訳：N1(3):彼 → 他 N2(1000):嘘 → 谎
- 訳文出力：
他的谎好.
- D 評価の原因：意味を理解できない。
普通は「他說的谎妙」と言う。

3.5 評価結果

3.2 節において選択した 200 文に対する，評価結果を表 1 にまとめる．

表 1 日中翻訳実験の結果

評価値	結果の割合
A	80 % (160/200)
B	7.5% (15/200)
C	5 % (10/200)
D	7.5% (15/200)

表 1 では，A 評価が 80 % となっており，作成した日中結合価パターン辞書は，単文の日中翻訳において大変有効であることが分かる．

すでに述べたように，評価の対象は使用頻度が高い結合価パターンなので，評価の結果から見ると，語義数の多い用言に対しても，結合語価パターン方式は有効な方法を明らかにしている．

また，パターン翻訳において，パターン化の基準は，対象とする言語ペアに依存すると考えられるが，この結果は，日英翻訳のために作成された日本語結合価パターンが，日中翻訳でもかなりの程度使用できることを示している．

これに対して，B，C と D の評価となった入力文（20% 40/200）は，日英翻訳で定められた日本語結合価パターンのカバー範囲が日中翻訳では適切でないこと，又は，結合価パターン方式の限界を示していることが考えられる．

4 考察

4.1 日本語結合価パターンのカバー範囲の問題

実験結果に基づき、日本語結合価パターンに適合した日本語文が、対応する中国語結合価パターンでは正しく翻訳できない場合について検討する。

ケース 1：意味的な制約条件のない名詞変数の問題

日英結合価パターンでは、意味的な制約条件の付与されていない変数がかなり存在する。これは、日英翻訳では、特定の格要素の意味属性で英語文型が決定される場合がかなり存在するためである。これに対して、日英翻訳で意味的な制約を不要とされていた変数の中にも、日中翻訳では、意味的な制約条件を付与すべき変数が存在する。これは、日英翻訳と日中翻訳では、訳し分けで重要な要素は同じでないことを示している。

3.3 節の C 評価の例 1 はこの例に当たる。この例を用いて、ケース 1 を説明する。
(パターン番号：302010-00、使う頻度：336 回)：

- 日本語結合価パターン：
でる：((N1 (-2671 *) ("が"))(N8 (*) ("に"))) でる

パターンの意味属性を表 2 に示す。

表 2 「パターン：でる」の意味属性調査表

属性番号	属性意味
*	全ての意味属性
-2671	暦日以外の意味属性

この例では、日本語結合価パターンに対して、意味属性の制約条件により、3 つの中国語結合価パターンを作成できる。作成例を表 3 に示す。

- 参考する日本文：
船員は甲板に出る。
- 参考する日中対訳文により、作成したパターン：
出現：N1 出現 在 N8

表3 「パターン：でる」に対して作成できる中国語結合価パターン

1. 出現： N1 出現 "在" N2
2. 体現： N1 体現 "在" N2
3. 出版： N1 出版 "在" N2

作成した中国語結合価パターンを用いて、クローズテストでは、良い翻訳品質を得られた。結果を以下に示す

- 変数の翻訳
船員 (*) → 船員
- 訳文の出力
船員出現在甲板。
- 評価値 → A

しかし、オープンテストでは、作成した結合価パターンを使用すると、文の愛昧さが出てくる。結果を以下の結果を示す。

- 日本語テスト文：
彼の性格が作品に出る。
- 変数の翻訳
性格 (*) → 性格
- 訳文出力
他的性格出現在作品。
- 評価値 → C

ケース1の問題で翻訳に失敗した入力文は、全体の約7%(14/200)に相当する。この問題を解決するには、現在の日英翻訳用の日本語結合価パターンの変数の意味分類体系をより詳細化する必要がある。

ケース 2：名詞の意味的な制約条件の粒度の問題

変数に対する意味的な制約条件が付与されている日本語結合価パターンでも，その条件が甘く，対応する中国語結合価パターンが複数存在するケースが多くある．試作した結合価パターン辞書では，そのうちの一つしか定義されておらず，誤った訳文が生成される．

3.3 節の D 評価の例文はこの例に当たる．この例を用いて，ケース 2 を説明する．
(パターン番号：203739-00、使う頻度：306 回)：
この例では，日本語結合価パターンに対して，意味属性の制約条件により，2 つの中国語結合価パターンを作成できる．作成例を表 4 に示す．

- 日本語パターン：
ひどい：((N1 (3 2055 1560 2422) ("が")))) ひどい

パターンの意味属性を表 4 に示す．

表 4 「パターン：ひどい」の意味属性調査表

属性番号	属性意味
3	主体
2055	出来事
1560	行為
2422	抽象的關係

この例では，日本語結合価パターンに対して，意味属性の制約条件により，4 つの中国語結合価パターンを作成できる．作成例を表 5 に示す．

表 5 「パターン：ひどい」に対して作成できる中国語結合価パターン

1. 厉害：N1 厉害
2. 無情：N1 無情

- 参考する日本文：
肩こりがひどい.
- 参考する日中対訳文により、作成したパターン：
厉害：N1 厉害

作成した中国語結合価パターンを用いて、クローズテストでは、良い翻訳品質を得られた。結果を以下に示す

- 変数の翻訳
N1(2055):肩こり → 肩酸
- 訳文出力
肩酸厉害
- 評価値 → A

しかし、オープンテストでは、作成した結合価パターンを使用すると、文の不適切さが出てくる。結果を以下の結果を示す。

- 日本語テスト文：
その子供の親はひどい。
- 変数の翻訳
N1(3):親 → 家长
- 訳文出力
那个孩子的家长厉害.
- 評価値 → D

ケース 2 の問題で翻訳に失敗した入力文は、全体の約 8%(16/200) に相当する。これらの問題を解決するには、現在の日英翻訳用の日本語結合価パターンの変数の意味属性の見直しが必要である。

ケース3：用言訳語選択の問題

日本語結合価パターンのカバー範囲が広く、複数の中国語結合価パターンが対応するような場合でも、一つの汎用的な中国語結合価パターンで済ますことのできる場合がある。例を以下に示す。

日本語結合価パターン：N1 が 綺麗

N1 の意味属性は：

(4人 468自然 534生物 760人工物 1002抽象物 2304自然現象 2564形状)

この例では、日本語の「綺麗」に対して、中国語の「美丽」、「好看」、「美观」の3つの述語が意味的に対応する。従って、これらの意味を訳し分けるには、表6のような3つのパターンが必要となる。

表6 「パターン：綺麗」に対して作成できる中国語結合価パターン

1. 美丽： N1 美丽
2. 好看： N1 好看
3. 美观： N1 美观

しかし、「好看」、「美观」は、それぞれ特殊な意味での「きれいさ」を表すのに対して、「美丽」は、より一般的な意味での「きれいさ」を表す。意味属性から言えば、(4人 468自然 534生物 760人工物 1002抽象物 2304自然現象 2564形状)

上記の意味属性に対して、「美丽」は全て対応できる。しかし、「好看」は(1002 抽象物 2304 自然現象)の部分の体言に対応すれば、不自然さがある。

例えば、(1002 抽象物)の恋情と対応すれば、「恋情好看」になると、とても不自然と感じる。

「美观」は(4人 468自然 534生物 1002抽象物 2304自然現象)の部分の体言に対応すれば、不適切さがある。例えば、(4人)の父と対応すれば、「父親美观」になると、意味を理解できないほどの不適切さがある。

そこで、日本語結合価パターンには、「美丽」を使用した中国語結合価パターンを対応させれば、意味の正しい中国語文が作成できる。すなわち、この日本語結合価パターンには、中国語結合価パターン「1. 美丽： N1 美丽」を対応させれば、正しい翻訳が可能となる。

実験結果によれば、このようなパターンの改良で正しい翻訳が可能になる入力文は約3%(6文/200文)である。

ケース4：不適切な名詞訳語の問題

入力文に対して適切な構造の中国語の訳文が作成されているときでも，変数部分（名詞）において不適切な訳語が選ばれていることがある．例を以下に示す．

日本語の名詞「木」は「樹」，「木头」，「椰子」などという3つの中国語訳語があり，意味的な用法として，それぞれ「樹木」，「材木」，「楽器」という意味属性を持つ．そこで，「木」の意味属性が決まることで，対応する訳語も決定できる．

ただし，変数化された日本語の単語に対して訳語が絞り込めない場合がある．文献 [4] によれば，このような語が日本語単語の全体の約 5% 存在する．この問題を解決するには，単語の意味属性体系を中国語単語の意味と整合するように見直すことが必要と考えられる．

4.2 結合価パターン方式の限界を超える問題

結合価パターンは、述部用言と格要素の意味的な関係を記述する枠組みであり、命題レベルでの、単文の意味を定義する方法として使用される。本節では、このような結合価パターンの限界を超える問題として、副詞的表現の翻訳と時制、相に関連する問題を取り上げる。

4.2.1 副詞の語順の問題

中国語の基本的な語順は「S(主語)+ Adv (副詞) + V(動詞) + Adj(形容詞) +O(目的語)」である。しかし、動詞、特有名詞、特殊の強調などにより、語順の例外がある。例えば、「明日学校に行く」を機械翻訳した場合、副詞「明日」は図3に示すように四つの場所に置くことができる。

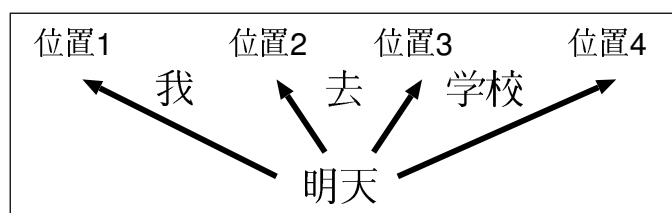


図3 「明日」の語順

位置1,位置2,位置4の意味は同じであるが,位置3の場合の意味は異なる。このような副詞の語順の問題を解決するには,副詞の要素も含めた文型パターン化が必要である。

4.2.2 時制・相により動詞が選択される問題

日本語では,通常,時制と相は助動詞によって表現される。これに対して,中国語の動詞は,動態動詞,静態動詞,結果動詞に分類され,動詞の種類によってこれらの情報が表される場合がある。特に,動態動詞は助動詞の補佐がないと,文の愛昧さ,および,不自然さが生じる。3.3節のB評価の例文はこの問題で正解とならなかったものである。この問題も前項と同様,結合価文法の枠組みを超える問題であり,これを解決するには,助動詞の要素も含めた文型パターン化が必要と考えられる。

5 おわりに

本研究では、使用頻度が高い日本語結合価パターン（200件）を対象に対応する中国語結合価パターンを作成し、日中機械翻訳における結合価パターン翻訳方式の可能性と問題点を検討した。その結果によれば、使用頻度が高く、意味的な訳し分けが必要と見られる日本語単文表現に対して、約80%は、正しい中国語訳文が得られることが分かった。従って、日英翻訳用に作成された結合価パターン対の日本語パターンは、日中翻訳のための結合価パターン対の作成においても有効だと言える。

また、翻訳誤りの分析によれば、誤りの大半は、日本語結合価パターンのカバー範囲の不適切さに起因していることが分かった。翻訳誤り20%のうちの15%は、日本語結合価パターンに適合した入力文が、必ずしも対応する中国語結合価パターンで訳すことはできず、意味によってより細かく訳し分けなければならないものであった。この問題を解決するには、中国語の表現構造に着目してそれに対応するように日本語結合価パターン自身を見直すこと、また、適合する日本文の範囲の適正化を図るため、日本語結合価パターン内の変数の意味的な制約条件を見直す必要のあることが分かった。

ところで、副詞の語順の問題や時制、相の問題で正しく訳せないものも5%程度存在するが、これらは、結合価文法の枠組みを超える問題であり、これらの問題を解決するには、副詞、助動詞などの表現要素を含むパターン辞書を開発する必要があると思われる。

本研究では、使用頻度の高い日本語結合価パターンの一部を対象に日中結合価パターン辞書を作成したが、日英翻訳用の結合価パターン辞書に収録された日本語結合価パターンは、日中翻訳でもかなり有効であることが分かったので、今後は、残された日本語結合価パターンに対しても中国語結合価パターンを作成し、実験的な改良を行うことにより、日中結合価パターン辞書を実現したい。

謝辞

この一年間で、本研究において指導していただいた知能情報工学講座、池原研究室の池原教授と村上助教授と徳久助手にお礼申し上げます。

そして、多大なる検討と助言をしてくださった計算機工学 C 研究室の方々に深く感謝します。

また、本研究で参考にした論文、本の著者の方々に礼申し上げます。

参考文献

- [1] 長尾ほか:自然言語処理,ISBN4-00-010355-5, 岩波書店,1996
- [2] 日本語語彙大系 ,NTT コミュニケーション科学研究所, 池原 悟ほか
- [3] 西山 七絵ほか :「単文文型パターン辞書の構築」, 言語処理学会第 11 回年次大会発表論文集 , pp.372-375.2005.
- [4] 展瑜ほか :「日中機械翻訳における名詞訳語の選択」, 言語処理学会第 9 回年次大会 C4-4 pp.334-337 2003.
- [5] 金出地真人ほか :「結合価文法による動詞と名詞の訳語選択能力の評価」 , 情報処理学会研究報告 2003-NL-153-16 pp.119-124 . 2003-01.
- [6] ALT/JE 関連
- [7] Munpyo Hong , Young-Kil Kim , Sang-Kyu Park , Young-Jik Lee
「Semi-Automatic Construction of Korean-Chinese Verb Patterns Based on Translation Equivalency」 Dep.of Speech/Language Technology Research , ETRI Korea , MLR 2004
- [8] Benjamin K. Tsou and Oi Yee Kwong:
「Evaluating Chinese-English Translation Systems for Personal Name Coverage」
Language Information Sciences Research Centre City University of Hong Kong
- [9] Chengqing Zong , Yujie Zhang:
「Approach to Spoken Chinese Paraphrasing Based on Feature Extraction」
Journal of Chinese Information Processing, Vol.17, No.6, pp.31-38.(in Chinese)