

概要

本研究では日中機械翻訳の手法の一つである結合価パターン翻訳方式を用いた日中機械翻訳システムを試作する。具体的には、IPAL 動詞 [1](計算機用日本語基本動詞、合計 955 語) を対象に日中結合価パターン辞書 (約 5 千件) を作成する [2] [3]。そして、作成した日中結合価パターン辞書を使用し、日中機械翻訳システムを試作する。更に、試作した翻訳システムを用いて、翻訳テストを行ない、翻訳能力を評価する。

オープンテストの結果により、翻訳文の全体には、評価値 A が 64(128/200) %、評価値 B が 16(32/200)% 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。

また、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 23%(46/200) 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、追加する必要があると考えている。

目次

1	はじめに	2
2	日中結合価パターン辞書	3
2.1	結合価パターン	3
2.2	日本語結合価パターン	3
2.3	IPAL 動詞	5
2.4	中国語結合価パターン	6
2.5	日中単語辞書	9
3	結合価パターンによる翻訳能力の評価	10
3.1	結合価パターンを用いた翻訳方法	11
3.2	試験用入力文	14
3.3	評価基準	15
3.4	評価の例	16
3.5	評価結果	24
4	考察	25
4.1	日本語結合価パターンのカバー範囲の問題	25
4.2	中国語結合価パターンの用言訳語の選択 (7%)	29
4.3	不適切な名詞訳語	31
4.4	結合価パターン方式の限界	32
5	おわりに	36

目次

1	日中翻訳実験の流れ	11
2	選択された日本語単文例	14
3	「明日」の語順	35

表目次

1	日中翻訳実験の結果	24
2	意味属性により作成できる中国語結合価パターン	27
3	意味属性により作成できる中国語結合価パターン	27
4	意味属性により作成できる中国語結合価パターン	29

1 はじめに

高品質な機械翻訳を目指し，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なわれている．日英翻訳では，既に網羅的なパターン辞書 [4] が開発され，パターン翻訳方式は有効であることを報告されている．しかし，日中翻訳において，大規模な文型パターン辞書を用いた翻訳方式の検討が少ないため，効果が不明確である [5]

そこで，本研究では日中機械翻訳を対象に文型パターン翻訳手法の一つである結合価パターン翻訳方式を用いた日中機械翻訳システムを試作する．そして，試作した翻訳システムを用いて，翻訳テストを行ない，翻訳能力を評価する．

ところで，結合価パターンによる翻訳方式は，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，使用頻度が高い IPAL 動詞を対象に，日中結合価パターン辞書を作成し，使用する．

オープンテストの結果により，翻訳文の全体には，評価値 A が 64(128/200) %，評価値 B が 16(32/200)% 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．

また，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 23%(46/200) 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，追加する必要があると考えている．

ところで，助動詞の問題や結果補語や時制，相の問題により正しく訳せないものも 26 文(全体の 13%) 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，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っている．今後は，改良を行うことにより，より精度が高い日中機械翻訳システムを実現したい．

2 日中結合価パターン辞書

2.1 結合価パターン

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る。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において、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。

本研究では、既に開発された日本語語彙大系を参考し、IPAL 動詞と対応する日本語結合価パターンを対象に、対応する中国語結合価パターンを作成する。

2.2 日本語結合価パターン

本研究では、使用する日本語結合価パターンは日本語語彙大系 [4] である。使用される日本語結合価パターンを以下に詳しく説明する。

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" N2 "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" N2 "を" 味わう

N1 の意味属性：(4 人)

N2 の意味属性：(838 食料 1253 感情)

- 例 3: 日本語結合価パターン：

N1 "が" N2 "に/で" 優勝する [状態 受身不可]

N1(3 主体 535 動物)

N2(1001 抽象物 1236 人間活動)

2.3 IPAL 動詞

自然言語処理において、膨大な情報を持つ電子化された辞書は非常に有効な資源の1つとなっている。計算機用辞書 IPAL もこうした日本語の電子化辞書の一つであり、情報処理振興事業協会 (IPA) により、自然言語処理に用いることを目的として作成されている。

この IPAL 辞書は言語学者によって、動詞、形容詞、名詞の各辞書ごとに順次作成されてきた辞書であり、各見出し語ごとに、計算機で利用できる情報(意味素性、述語素、複合語など)が極めて詳細に記述されているという特徴がある。そのため、これまで多くの研究に利用されてきた。

また、本研究では IPAL 動詞辞書(計算機用日本語基本動詞辞書)を使用する。IPAL 動詞辞書は日本語の動詞のうちで語彙体系上ならびに使用頻度上重要であると考えられる基本的な和語動詞 861 語とサ変動詞 94 語が取り上げられている。意味および統語的特徴に基づいて下位区分し、それを1つの単位として、意味、形態、統語、文法的カテゴリ、慣用表現などに関わる情報をできるだけ詳細に記述したものである。

日本語の動詞「くう」に関する例を以下に示す。

"くう","0 0 1","0 0 3","0 0 5","重要動詞 X 0 1 3","相手の勢力 範囲などを侵す。","越える","凌ぐ","","(競争・攻防・勝敗),(所有・取得)","(2 . 3 5 3),(2 . 3 7 0)","受動・能動","2 9 5 a","動作(動き),(抽象的)関係","食(く)う,喰(く)う","五段","k u w","","他","くえる","","","","","","","","大物ぐい","N 1 ガ N 2 ヲ","子役が 主役を 食ってしまった。","新人候補が 現役候補を食った。","","O 2 a , A","ガ","H U M / O R G","子役、平幕 / スーパーマーケット","ヲ","H U M / O R G / A B S","主演女優、横綱 / 小売店、地元商店街 / 相手の地盤","","","","","","","","","","","二使役","直受、間受、尊敬","ヲ","二","","","能動","未来","経験・完了","テシマウ、カケル","命令","意志・勧誘","タイ・タガル、ナサイ、ナ(禁止)","3 a"

2.4 中国語結合価パターン

本研究では，実験用中国語結合価パターン辞書が存在しないため，予め，人手で作成する．作成した 4903 件のパターンは中国語結合価パターン辞書として使用する．以下に示す 4 つのステップにより，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．

1. 対象とする日本語結合価パターンの選択

日本語語彙大系 [4] から，IPAL 動詞に対応する日本語結合価パターン（4,903 件）を選択する．

2. 日本文の選択

上記で選択された各日本語結合価パターンに対して，日本語単文集 [6] から，適合する日本文を 2 文選択し，選択した日本文は更に 4 つの条件を満たすように修正する．

- (1) 副詞句を含まない．
- (2) 形容詞句と名詞句は対象内とする．
- (3) 全ての格要素は結合価パターンに含まれる．
- (4) 用言の終止形で終る．

3. 中国語訳文の付与

選択した日本文に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 2．と 3．で得られた日中対訳文を参考に，1．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（4,903 件）を作成する．

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（4,903 件）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" N2 "を" 開ける
- 参考した日本語例文と中国語訳文
日本語例文 1 : 生徒は引出しをあける。
中国語訳文 1 : 学生打開抽屉。
日本語例文 2 : 彼は窓を開ける。
中国語訳文 2 : 他打開窗户。
- 中国語結合価パターン：
N1 打開 N2
- 体言の意味属性：
N1 : (3 主体) , N2 : (533 具体物 389 施設)

例 2 :

- 日本語結合価パターン：
N1 "が" N2 "を" 食う
- 参考した日本語例文と中国語訳文
日本語例文 1 : 大型冷蔵庫は電気を食う。
中国語訳文 1 : 大型冰箱需要使用电。
日本語例文 2 : 車はガソリンを食う。
中国語訳文 2 : 车需要使用汽油。
- 中国語結合価パターン：
N1 需要使用 N2
- 体言の意味属性：
N1 : (962 機械 986 乗り物) N2 : (2335 電気 793 燃料)

例 3 :

- 日本語結合価パターン :

N1 "が" N2 "に" 潜る

- 参考した日本語例文と中国語訳文

日本語例文 1 : 彼は床下に潜る .

中国語訳文 1 : 他藏在床下 .

日本語例文 2 : その組織は地下に潜る .

中国語訳文 2 : 那个团体藏在地下 .

- 中国語結合価パターン :

N1 藏在 N2 "下"

- 体言の意味属性 :

N1 : (3 主体) N2 : (2636 下)

2.5 日中単語辞書

結合価パターンを利用するために、体言の意味体系が必要である。本研究では、体言の意味属性を付与した日中単語辞典(約 1,200 語)を手作業で作成する [7]。

例を以下に示す。

日本語の名詞「木」は「樹」、「木头」、「椰子」という3つの中国語訳語がある。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、対応する訳語が決定できる。

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，IPAL 動詞を対象に作られた日中結合価パターン対の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

結合価パターンによる翻訳能力を評価するため，図 1 に示すような流れで機械翻訳システムを試作する．

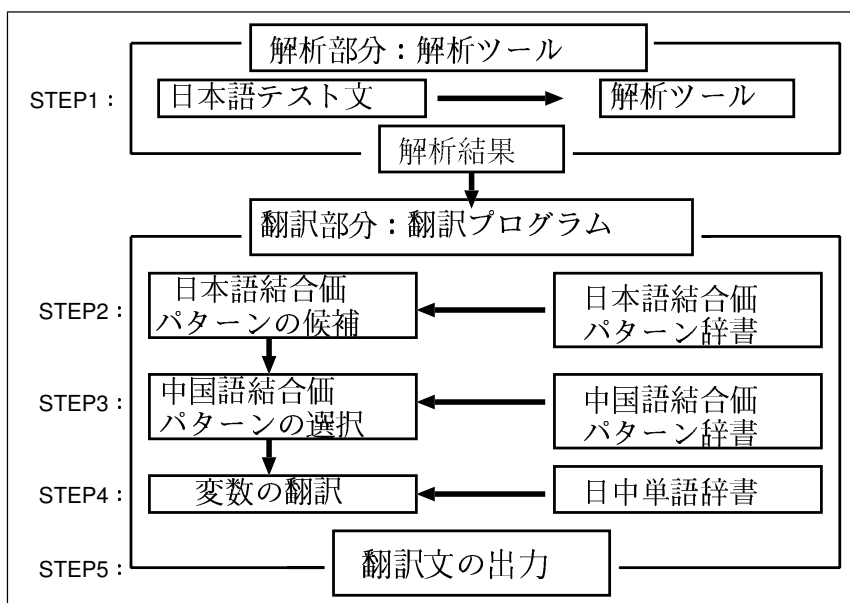


図 1 日中翻訳実験の流れ

翻訳システムは 2 つの部分で分かれている．解析部分と訳文生成部分である．

具体的な翻訳手順と翻訳例を以下に示す．

解析部分

図 1 に示したステップ 1 において，入力された日本語テスト文を解析ツールを通して，形態素解析を行なう．また，入力文は日本語単文集 [6] 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．入力文と解析の例を以下に示す．

- 入力文：中国は日本にパンダを贈る．

形態素解析結果：

- 1.) 中国 (367 公共機関,385 国家,405 学校,463 領土,30 独立国,98 大学・高専)/は (7530)
- 2.) 日本 (385 国家,463 領土,30 独立国)/に (7430)

3.) パンダ (537 獣)/を (7430)

4.) 贈る (2386,19 所有的移動)/。 ([P]0110)

訳文生成部分

上記解析部分で分析した結果を翻訳プログラムに入力する。翻訳プログラムは日中結合価パターン辞書と日中単語辞書を使用し、以下の手順で、各日本語入力文に対する中国語訳文を作成する。

(1) 日本語結合価パターンの選択

図1に示したステップ2では、テスト文と適合する日本語結合価パターンを調べる。同じ用言であり、さらに、入力文が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定する。また、適合した日本語結合価パターンとパターンIDを出力する。上記に入力したテスト文に対する出力結果を以下に示す。

- 日本語結合価パターン：

N1 が N2 を N3 に/へ 贈る

N1(3 主体) N2(533 具体物 1001 抽象物) N3(3 主体)

- パターンID：301429-00

(2) 中国語結合価パターンの出力

図1に示したステップ3では、同じパターンIDの中国語パターンを出力する。

- 中国語結合価パターン：

N1 送給 N3 N2

(3) 単語の翻訳

図1に示したステップ4では、適合した変数の値(日本語単語)に対して、日中単語辞書から、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。日中単語辞書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。

- 訳語の翻訳：

中国 (385 国家)→ 中国

日本 (385 国家)→ 日本

パンダ (537 獣)→ 熊猫

(4) 訳文の生成

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．訳文の生成結果を以下に示す．

- 中国語訳文：中国送給日本熊猫．

次の節から，日中翻訳実験で，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 [7] 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. 2.4 節において使用した単文ではない。
2. 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。
3. 日本語単文が用言の終止形で終る。
4. 形容詞句と名詞句は対象内、副詞句を含まない。
5. 慣用表現を含まない単文。

選択した入力文の例を図 2 に示す。

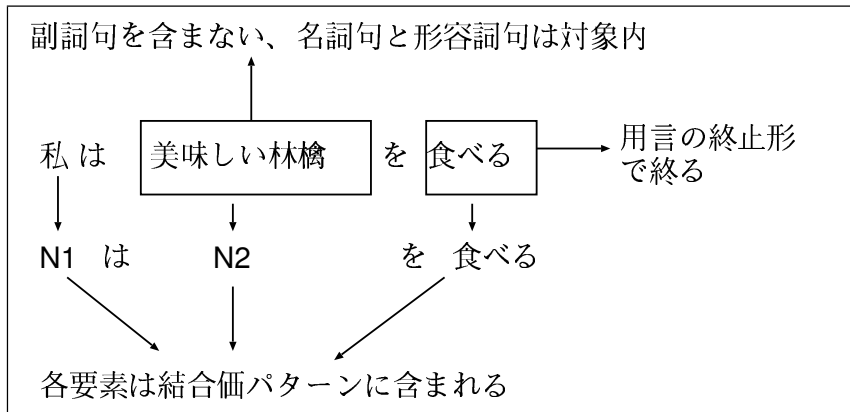


図 2 選択された日本語単文例

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) 曖昧なところがあるが、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 : 風は南西から北へ吹く
- 使用された結合値パターン対 :
日本語パターン : (" 吹く")
N1 "が" N2 "から, より" N3 "に, へ" 吹く
中国語パターン: ("吹")
N1 "从" N2 吹向 N3
N1 の意味属性 : (2373 風)
N2 の意味属性 : (388 場所)
N3 の意味属性 : (388 場所)
- 変数の翻訳 :
風 → 风 , 南西 → 西南 , 北 → 北方
- 訳文出力 : 风从西南吹向北方 .

例 2:

- テスト文 :
私はラケットでボールを打つ
- 使用された結合値パターン対 :
日本語パターン : ("打つ")
N1 "が" N2 "を" N3 "で" 打つ
中国語パターン: ("打")
N1 "用" N3 打 N2
N1 の意味属性 : (4 人)
N2 の意味属性 :
(552 動物 (部分) 921 遊び道具・運動具 2020 打ち 533 具体物)

N3 の意味属性：(-955 棒 863 建造物)

- 変数の翻訳：

私 → 我，ボール → 球，ラケット → 球拍

- 訳文出力：我用球拍打球.

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")
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：三杯の水→三杯的水
- 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：
私は叫び声を聞く
- 使用された結合価パターン対：
日本語パターン：("聞く")
N1 "が" N2 "を" N3 "から" 聞く
中国語パターン：("听")
N1 "从" N3 听 N2
N1 の意味属性 (3 主体 535 動物)
N2 の意味属性 (1000 抽象)
N3 の意味属性 (3 主体)
- 変数の翻訳：私 → 我，叫び声 → 叫声
- 訳文出力：我听叫声.

- B 評価の原因：
結果補語がないので，文の全体は不自然である．通常「我听到/着叫声」である．

例 3:

- テスト文：
火事は四方に広がる．
- 使用された結合価パターン対：
日本語パターン：("広がる")
N1 "が" N2 "に" 広がる
中国語パターン：("蔓延")
N1 蔓延 "到" N2
N1 の意味属性 (2507 状態 1239 感覚 1253 感情 2416 病気 2056 災難 1218 流行 695 枝 2355 におい 2356 味)
N2 の意味属性 (3 主体 388 場所 2610 場 552 動物 (部分))
- 変数の翻訳：火 → 火，四方 → 四方
- 訳文出力：火蔓延到四方．
- B 評価の原因：名詞訳語選択問題である．文の全体は不自然であり，通常「火蔓延到各处」である．

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：(3 主体)
N8 の意味属性：(1120 目録,1057 創作物,970 通信機器など)
- 変数値の翻訳：彼の性格 → 他的性格 , 作品 → 作品
- 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：
彼女が日本酒を冷やす.
- 使用された日中結合価パターン対：
日本語パターン：("冷やす")
N1 "が" N2 "を" 冷やす
中国語パターン：("冷却")
N1 冷却 N2
N1 の意味属性：(3 主体 760 人工物)
N2 の意味属性：(全ての意味属性)
- 変数値の翻訳：彼女 → 她 , 日本酒 → 日本酒
- 訳文出力：：她冷却日本酒 .
- C 評価の原因：意味を理解しにくい .
普通は「她冰鎮日本酒」と言う .

例 3 :

- テスト文：
仏教はアジアに広まる。
- 使用された日中結合価パターン対：
日本語パターン：("広まる")
N1 "が" N2 "に,へ" 広まる
中国語パターン：("扩散")
N1 扩散 "到" N2
N1 の意味属性：(すべての意味属性)
N2 の意味属性：(388 場所 2610 場)
- 変数値の翻訳：仏教 → 佛教 , アジア → 亚洲
- 訳文出力：：佛教扩散到亚洲。
- C 評価の原因：意味を理解しにくい。
普通は「佛教发展到亚洲」と言う。

3.4.4 D 評価の例

例 1:

- テスト文：
彼は出世の足を引っ張る。
- 使用された日中結合価パターン対：
日本語パターン：("引っ張る")
N1 "が" N2 "を" 引っ張る
中国語パターン：("延长")
N1 延长 N2
N1 の意味属性：(人)
N2 の意味属性：(すべての意味属性)
- 変数値の翻訳：彼 → 他，出世の足 → 出人头地
- 訳文出力：他延长出人头地。
- D 評価の原因：意味を理解できない。
普通は「他影响出人头地。」と言う。

例 2:

- テスト文：
彼はタバコをやめる。
- 使用された結合価パターン対：
日本語パターン：("やめる")
N1 "が" N2 "を" やめる
中国語パターン：("停止")
N1 "停止" N2
N1 の意味属性：(4 人)
N2 の意味属性：(862 たばこ)
- 変数値の翻訳：彼 → 他，タバコ → 烟
- 訳文出力：他停止烟。
- D 評価の原因：意味を理解できない。
通常「他禁烟。」である。

例 3:

- テスト文：
ここから線路が西方へ分かれている。
- 使用された結合価パターン対：
日本語パターン：("分かれる")
N1 "が" N2 "から,より" N3 "に,へ" 分かれる
中国語パターン：("分散")
N1 "从" N2 "向" N3 分散
N1 の意味属性：(417 交通路 2567 線)
N2 の意味属性：(全ての意味属性)
N3 の意味属性：(全ての意味属性)
- 変数値の翻訳:線路 → 线路 , ここ → 这 , 西方 → 西方
- 訳文出力：线路从这向西方分散。
- D 評価の原因：意味を理解できない。
通常「线路从这向西方分支」である。

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	64%(128/200)
B	16%(32/200)
C	13%(26/200)
D	7%(14/200)

表 1 では，A 評価が 64% 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った．

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 36%(72/200) 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

オープンテストの実験結果に基づき，評価値 B 以下の 72 文の原因を調査する．

4.1 日本語結合価パターンのカバー範囲の問題

本研究で使用した日本語語彙大系の結合価パターンは日英翻訳のために作ったものである．日中機械翻訳において，日本語結合価パターンのカバー範囲の問題を生じる．この問題により，翻訳に失敗した入力文は，全体の 23%(46 文/200 文)である．この問題は 4.1.1 節と 4.1.2 節で解説した 2 つのケースがある．これらの問題を解決するには，現在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り，また，その際，必要に応じて，名詞の意味分類体系をより詳細化する必要があると考えている．

4.1.1 意味的な制約条件がない名詞変数 (12.5%)

日本語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数多く存在する．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多く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．評価値 B 以下の 72 文中の 25 文 (全体の 12.5%) はこの問題に起因する．25 文中では，B 評価 9 文，C 評価 8 文，D 評価 8 文だった，また，例を以下に示す．

例 1

日本語結合価パターン：

N1(3 760) "が" N2(*) "を" 冷やす

中国語結合価パターン：

N1 冷却 N2

体言の意味属性：

N1(3 主体 760 人工物)，N2(全ての意味属性)

入力文：彼女が日本酒を冷やす．

翻訳文：她冷却日本酒．

正しい翻訳：她冰鎮日本酒.

原因：名詞の意味的な制約条件が広すぎ、用言「冷却」が対応できない。この原因で C 評価になった。

例 2

日本語結合価パターン：

N1 (*) "が" N2(388 2610) "に,へ" 広まる

中国語結合価パターン：

N1 扩散 "到" N2

体言の意味属性：

N1(すべての意味属性) N2(388 場所 2610 場)

入力文：仏教はアジアに広まる。

翻訳文：佛教扩散到亚洲。

正しい翻訳：佛教发展到亚洲。

原因：名詞の意味的な制約条件が広すぎ。用言「扩散」が対応できない。この原因で C 評価になった。

4.1.2 名詞の意味的な制約条件の粒度の問題 (10.5%)

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、その条件が広く、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、そのうちのひとつしか定義されておらず、誤った訳文が生成される。評価値 B 以下の 72 文中の 21 文 (全体の 10.5%) はこの問題に起因する。21 文中では、B 評価 11 文、C 評価 7 文、D 評価 3 文だった。また、3.3 節の C と D 評価の例文は、この例に相当する。

例 1

使用したテスト文：彼の性格が作品に出る。

日本語パターン：("でる")

N1 "が" N2 "に" でる

N1 の意味属性：(主体)

N2 の意味属性：(1120 目録,1057 創作物,970 通信機器など)

この日本語結合価パターンに対して、表 2 に示す 3 つの中国語結合価パターンが対応

する .

表 2 意味属性により作成できる中国語結合価パターン

1 . 出現 (現れる) : N1 出現 "在" N2(-920 出版物)
2 . 体現 (性格を表現する) : N1 体現 "在" N2 (-920 出版物)
3 . 出版 (出版する) : N1 出版 "在" N2 (920 出版物)

実際に使用した中国語結合価パターン : ("出現")

N1 出現 "在" N2

変数値の翻訳 : 性格 → 性格 , 作品 → 作品

中国語で「他の性格出現在作品 .」と言わず , 意味が理解し難くなるため , C 評価になった . 通常「他の性格体現在作品 .」である .

例 2

使用したテスト文 : 彼はタバコをやめる .

日本語パターン : ("やめる")

N1 "が" N2 "を" やめる

N1 の意味属性 : (4 人)

N2 の意味属性 : (862 たばこ)

日本語結合価パターンに対して , 2 つの中国語結合価パターンが対応する具体的な例を表 3 に示す .

表 3 意味属性により作成できる中国語結合価パターン

1 . 停止 (停止する) : N1 停止 N2(862 たばこ)
2 . 禁 (禁止する) : N1 禁 N2(862 たばこ)

実際に使した中国語結合価パターン：("停止")

N1 "停止" N2

N1 の意味属性：(4 人)

N2 の意味属性：(862 たばこ)

変数値の翻訳:彼 → 他 , タバコ → 烟

「他停止烟 .」と翻訳され,意味が理解出来ず ,D 評価になった . 通常「他禁烟 .」である .

4.2 中国語結合価パターンの用言訳語の選択 (7%)

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にまとめることができる．汎用的な中国語結合価パターンが存在しないため，翻訳に失敗した入力文は 200 文中 14 文 (全体の 7%) だった．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：

(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 4 に示す3つのパターンが必要となる．

表 4 意味属性により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」を表すため，上記の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．

翻訳実験においての例

日本語結合価パターン：

N1 (3) "が" N3 (517) "に" N2 ("水") "を" 引く

この例では，日本語の「引く」に対して，中国語の「引入」，「引进」の2つの述語が意味的に対応する．

使用した中国語結合価パターン：

N1 "往" N3 "里" 引进 N2

体言の意味属性：

N1 (3 主体), N3 (517 耕地・牧場), N2 ("水")

入力文：彼は田に水を引く。

翻訳文:他往田里引进水。

正しい翻訳：他往田里引入水。

原因：翻訳文に違和感があるため，B 評価になった。用言の「引进」の目的語は知識類のことが多い。それに対して，「引入」は，より一般的な意味である。

4.3 不適切な名詞訳語

翻訳テストでは，入力文に対して適切な構造の中国語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．

例 1

日本語結合価パターン：

N1 (3) "が" N2(2 1009) "を" N8 (2 2610) "に,へ" 纏める

中国語結合価パターン：

N1 把 N2 归纳成 N8

入力文：私が全員の考えを一つに纏める．

翻訳文：我把全员的想法归纳成一个．

正しい翻訳：我把全员的想法归纳成一体．

原因：訳語の選択問題である．用言「归纳」に対して，名詞の訳語「一个」はより自然と感ずる．翻訳された文が理解しにくいいため，C 評価になった．

例 2

日本語結合価パターン：

N1 (2507 1239 1253 2416 2056 1218 695 2355 2356) "が" N5 (3 388 2610 552) "に" "へ" "まで" 広がる

中国語結合価パターン：

N1 蔓延到 N5

入力文：火事は四方に広がる．

翻訳文：火蔓延到四方。

正しい翻訳：火蔓延到各处。

原因：訳語の選択問題である．用言「蔓延」に対して，名詞の訳語「各处」はより自然と感ずる．翻訳された文が不自然のため，B 評価になった．

また，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は文献 [7] によれば，日本語単語の全体の約 15% 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しようと考えられる．

4.4 結合価パターン方式の限界

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、命題レベルにおいて、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、時制、相に関連する問題、状態補語の問題と副詞的表現の翻訳問題を取り上げる。

4.4.1 時制・相により動詞が選択される問題 (2.5%)

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。評価値 B 以下の 72 文中の 5 文 (全体の 2.5%) はこの問題に起因する。以下に例を示す。

例 1 :

- テスト文 :
私は叫び声を聞く
- 使用された結合価パターン対 :
日本語パターン : ("聞く")
N1 "が" N2 "を" N3 "から" 聞く
中国語パターン : ("听")
N1 从 N3 听 N2
N1 の意味属性 (3 主体 535 動物)
N2 の意味属性 (1000 抽象)
N3 の意味属性 (3 主体)
- 変数の翻訳 : 私 → 我, 叫び声 → 叫声
- 訳文出力 : 我听叫声 .
- B 評価の原因 : 状態を表す助動詞がないので, 文の全体は不自然である. 通常「我听到叫声 .」または「我听着叫声 .」である .

例 2:

- 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")
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：三杯の水→三杯的水
- 訳文出力：
喝 3 杯的水.
- B 評価の原因：
助動詞がないので，文の全体は不自然である. 普段は「喝了 3 杯的水。」と言う.

この問題は，結合価文法の枠組みを超える問題であり，この問題を解決するには，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．

4.4.2 状態補語問題 (3.5%)

日本語では、通常、状態補語を使わない。例えば、「例 1：島に 虎がいる。」や「例 2：家に テレビがある」などに対して、「島の上に虎がいる」と「家の中にテレビがある」と言わなくても、誤解を生じない。

これに対して、中国語では状態補語がよく使われる。例文に対して、通常、「例 1：在島上 有老虎」と「例 2：在家里 有电视」という。しかし、状態補語は名詞により変わり、動詞の受身状態も関係があるので、一つの結合価パターンに定義するのは困難である。評価値 B 以下の 72 文中の 7 文 (全体の 3.5%) がこの問題に起因する。例を以下に示す。

- テスト文：
彼は壁に大穴を開ける
- 使用された日中結合価パターン対：
日本語パターン：("明ける")
N1 "が" N2 "に" N3 "を" 明ける
中国語パターン：("垂")
N1 "在" N2 垂 N3
N1 の意味属性：(3 主体 962 機械)
N2 の意味属性：(533 具体物)
N3 の意味属性：(2574 穴)
- 変数の翻訳：彼→他，壁→墙，大穴→洞
- 訳文出力：他在墙垂洞。
- C 評価の原因：
状態補語がないため、意味が理解しにくい。普通「他在墙上垂洞」と言う。

この問題に対して、体言の意味属性をより詳細化分類する必要がある。また、体言間の係り関係や体言と用言の受身関係などの情報により、動詞状態を明確化する必要があると考えている。

4.4.3 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」は以下のように4つの場所に置くことができる。

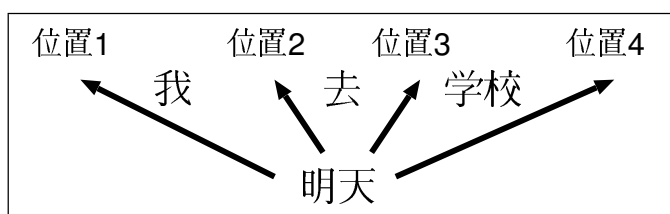


図3 「明日」の語順

(位置1) 明天 我去学校。

(位置2) 我 明天 去学校。

(位置3) 我去 明天 学校。

(位置4) 我去学校 明天。

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。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。

また，副詞の語順によって，助動詞が変わる場合もある。例を以下に示す。なお，アンダーラインは助動詞の印である。

日本語：私は速く走れる。

訳語：私→我，速く→飞快，走れる→跑

中国語の対訳文：

(1) 我跑得飞快。

(2) 我飞快地跑。

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。

5 おわりに

本研究では、IPAL 動詞を対象にした日中結合価パターン辞書を使用し、日中機械翻訳システムを試作した。また、作成した翻訳システムを用いて、翻訳実験を行ない、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。

オープンテストによれば、200 文の入力文に対して、128 文は正しい中国語訳文が得られることが分かった。その結果から、開発した日中結合価パターン辞書は日中機械翻訳システムの構築に効果があることが分かった。

翻訳誤りの分析結果によって、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 72 文のうちの 46 文 (全体の 23%) 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ことが必要であると考えられる。

また、助動詞の問題や結果補語や時制、相の問題により正しく訳せないものも 26 文 (全体の 13%) 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、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。今後は、改良を行うことにより、より精度が高い日中機械翻訳システムを実現したい。

また、体言の意味属性を付与した日中辞典が存在しないため、本研究では、手作業で意味属性を考えた日中辞典を作成した。しかし、登録した単語の数が少ないため、より規模が大きい翻訳テストを実現できなかった。今後は、意味属性を付与した日中辞典を電子化したいと考えている。

謝辞

この三年間で、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] 情報処理振興事業協会. 計算機用日本語基本動詞辞書, 1999 .
- [2] 楊 鵬ほか : 「日中機械翻訳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 回年次大会発表論文集, pp.79-82, 2007.
- [3] 楊 鵬ほか : 「結合価パターンを用いた日中機械翻訳システムの構築」情報処理学会研究報告会, pp.121-126, 2008.
- [4] 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
- [5] 長尾ほか:自然言語処理, ISBM4-00-010355-5, 岩波書店, 1996
- [6] 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.
- [7] 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.
- [8] 楊 鵬ほか : 「結合価パターンを用いた日中機械翻訳方式の検討」, 言語処理学会第 12 回年次大会発表論文集, pp.264-267, 2006.
- [9] 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.
- [10] ALT/JE 関連
- [11] 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
- [12] 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
- [13] 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)