

概要

等価的類推思考の原理により，重文・複文の日英翻訳のために文型パターン辞書を構築している．この翻訳方式は，日英の対訳のとれた文を用い，日英において言語表現を言い替えても意味の通る部分を，*VP*（動詞句），*NP*（名詞句），*V*（動詞），*N*（名詞）などのように変数で表し，言い替えると意味の通らない部分を字面で残しており，日英の両パターンを合わせて文型パターン対としている．また，言い替え可能な要素を線形要素と呼び，言い替え不可能な要素を非線形要素と呼ぶ．そして，変数でバインドされた部分表現ごとに翻訳を行い，英語のパターンにそれぞれの部分表現の対訳を代入することにより，英文を生成する．この翻訳方式を局所翻訳という．そこで「線形要素の翻訳が，代入された日本語だけで可能である」という仮説を立てている．しかし，部分表現ごとに翻訳することについて，一部の情報のみで英訳が生成できるのかという，局所翻訳の可能性の程度はまだ分かっていない．

本研究では，部分表現の中でも動詞句についての検証を行う．具体的に，次の2点についての検証を行う．

- 局所翻訳の可能性の検証
- 結合価パターンの有効性検証

結合価パターンは，単文の翻訳に有効であることが分かっている．単文と動詞句は良く似ており，単文の主語を除いたものを動詞句として見なすこともできる．そのため，本研究では結合価パターンが動詞句の翻訳にどの程度有効であるかを検証する．

結合価パターンの有効性の検証については，結合価パターンのみでの検証を行うため，本研究にて，翻訳プロトタイプ「queen」を作成する．局所翻訳の可能性の検証については，queen と，参考として，一般の翻訳ソフトである参考1（「ALT-J/E」），参考2（「翻訳の王様」）にて検証を行う．具体的な検証方法は次の通りである．まず，重文・複文の約12万文対の中で，日英の対訳関係の見出された約7万件の動詞句対を対象とする．ただし，英語の動詞句を構成する単語数ごとに分類し，その中で2単語から14単語までのものを，それぞれランダムで25句ずつ取り出した計325句対とする．次にこの325句対に対し，3つのシステムでそれぞれ翻訳実験を行い，訳文の比較を行う．

検証の結果として，局所翻訳の可能性は現段階で約70%～76%であることが分かった．また，queen のみの結果では，再現率が89.5%，適合率が75.9%と，結合価パターンのみを使用しただけの翻訳システムにおいては，良い結果が得られたと言える．よって，動

詞句の翻訳に結合価パターンは有効であることが示された。

今後、queen のパターン照合条件の修正、結合価パターンに存在しない動詞の追加、意味属性制約への柔軟な対処を行い、結合価パターンの範囲外である、副詞についても実装し、格要素内の名詞を名詞句として認識することができれば、更に精度があがると考えられる。

目次

1	はじめに	5
2	研究の背景	6
2.1	結合価文法	6
2.1.1	一般名詞意味属性体系	6
2.1.2	結合価パターン	7
2.2	文型パターン	7
2.3	本研究の目的	8
3	結合価パターンを用いた動詞句翻訳	10
3.1	翻訳用知識ベース(結合価パターンの変形)	10
3.1.1	結合価パターンの句形成	10
3.2	補語の削除	11
3.3	翻訳手順	13
4	実験	15
4.1	実験の目的	15
4.2	翻訳対象	15
4.3	訳文の評価	15
4.3.1	訳文の評価値と評価基準	15
4.3.2	評価方法	16
4.3.3	評価の例	16
4.4	実験結果	17
5	考察	19
5.1	局所翻訳の可能性の検証について	19
5.2	結合価パターンの有効性の検証について	19
5.3	諸般の問題	24
5.3.1	補語の削除の問題	24
5.3.2	対訳英語が1単語で構成される場合	24
6	おわりに	27

目 次

1	一般名詞意味属性体系の一部	6
2	各単語数でのスコア	19

表 目 次

1	結合価パターンの一般表現と慣用表現の一例	7
2	3つのシステムの翻訳結果	17
3	queen, 参考1, 参考2の再現率 R と適合率 P	18
4	理想解の構成単語が1の時の実験結果	25

1 はじめに

要素合成法を基本とした従来の機械翻訳方式の限界を突破する方法として、非線形な言語表現の構造を意味のまとまる単位にパターン化した文型パターン翻訳方式が提案されている [1]。この方式では、線形要素は変数で表し、変数に対する訳出表現を組み合わせることで全体の英文を作成する。この翻訳方式を局所翻訳と言う。そこで、「線形要素の翻訳が、代入された日本語だけで可能である」という仮説を立てている。しかし、部分表現ごとに翻訳することについて、一部の情報のみで英訳が生成できるのかという、局所翻訳の可能性の程度はまだ分かっていない。

本研究では、変数の中でも動詞句の翻訳に焦点を当て、動詞句の翻訳可能性について調査する。調査のねらいは、次の2点である。

1. 局所翻訳の可能性：動詞句の翻訳において、上述の仮説の検証
2. 結合価パターンの有効性：単文の翻訳では、動詞句の訳語選択に結合価パターンが有効である [2] が、単文と良く似た動詞句の翻訳についての有効性の検証

まず、2については、結合価パターン辞書のみでの検証を行うため、本研究で翻訳プロトタイプ「queen」を作成し、検証を行う。

1については、queen と、参考として、一般の翻訳ソフトである、参考1 (ALT-J/E)、参考2 (「翻訳の王様」) を用いて検証を行う。具体的な検証方法は次の通りである。まず、重文・複文の約12万文対の中で、日英の対訳関係の見出された約7万件の動詞句対を対象とする。ただし、英語の動詞句を構成する単語数ごとに分類し、その中で2単語から14単語までのものを、それぞれランダムで同じ句数ずつ取り出すこととする。

2 研究の背景

パターン翻訳において、現在、単文と、重文・複文についてパターン辞書が構築されている。単文についてのパターン翻訳は、多段翻訳方式に基づき、日本語語彙大系 [3] にて構築した結合価文法に従って翻訳を行っている。結合価パターンに記述された名詞変数 N の局所翻訳は、[4] にて有効性があることが分かっている。重文・複文については、等価的類推思考の原理により、最近、[1] にて重文・複文の文型パターン（以下、文型パターンと称す）が構築された。しかし、文型パターンに記述された変数の局所翻訳の可能性は、まだ示されていない。

2.1 結合価文法

2.1.1 一般名詞意味属性体系

一般名詞意味属性体系とは、単語の意味により体系的に分類、整理した語彙集である。本研究で使用する一般名詞意味属性体系は、約 40 万語の名詞を、最大 12 段（図 1 中の*1~12、名詞の属性の深さを表す）の木構造を構成する 2,710 の意味属性に分類している。また、一般名詞意味属性体系は、木構造を基本構成とし、上位の属性を持つ名詞は、木構造において自分より下位の名詞の意味属性を包含する性質がある。

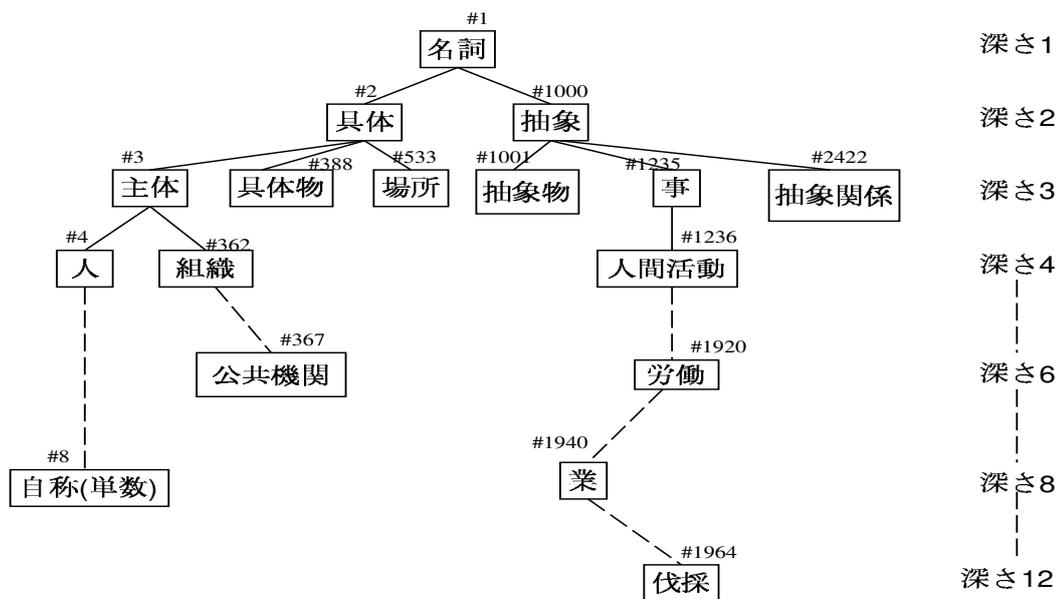


図 1: 一般名詞意味属性体系の一部

2.1.2 結合価パターン

結合価パターンは、用言と格要素（名詞＋格助詞）の意味的關係を記述したものである。用言と格要素間に意味属性による制約を付ける事で訳語選択の精度を高める。本研究では、結合価パターンの名詞の意味属性を調べるために、一般名詞意味属性体系（図1）を用いる。日本語語彙大系に登録してある結合価パターンにある名詞は、一般名詞意味属性体系によって意味属性（図1中の#1～1964下の単語。#の数字は意味属性番号）を付けている。上位の意味属性が下位の意味属性を包含する事と名詞の持つ意味属性を利用し、結合価パターンは動詞の持つ意味を規定することができる。

本研究では、[3]にある「構文意味辞書」の結合価パターンを基本にする。結合価パターンは、一般表現と慣用表現を合わせ、約15,000件の日本語文型パターンにまとめている。結合価パターンの例を表1に示す。

表 1: 結合価パターンの一般表現と慣用表現の一例

一般	$N1(3 \text{ 主体})$ が $N2(3 \text{ 主体})$ を 援助する	$N1 \text{ help } N2$
慣用	$N1(838 \text{ 食料})$ が 口に 合う	$N1 \text{ taste good}$

なお N とは名詞変数であり、括弧内は名詞の意味属性番号と意味属性である。

また、一般表現は、用言と一つ以上の格要素で規定する。慣用表現は、名詞が意味属性でなく直接単語の字面で規定され、主に特殊な表現に用いる。

2.2 文型パターン

[1]にて作成した文型パターン辞書の一例を以下に示す。

日本文：大勢の前で話すときは多くの人があがるものだ。

対訳英語文：Many people have stage fright when they have to give a speech in front of a large audience.

日本語パターン： $VP1$ ときは $NP2$ が $V3$ ものだ。

対訳英語パターン： $NP2 V3$ when $N3$ have to $VP1$.

日本文と対訳英語文において、それぞれ置き換え可能な線形要素の部分を変数にしたパターン対が作られている。日本語パターン、対訳英語パターンの変数 $VP1$ には「大勢の前で話す」が、 $NP2$ には「多くの人」が、 $V3$ には「あがる」が適合している。対訳英

語パターンを用いて英文を生成するためには，それぞれの変数に適合する日本文の対訳を生成し，代入する必要がある．動詞は辞書でカバーできると考えられるが，動詞句や名詞句は別ルーチンを用いて翻訳する必要があると思われる．その中で，本研究では動詞句を対象とする．また，名詞句については，今年度の卒業研究で，神野絵理さんが研究を行っている．

2.3 本研究の目的

先ほどの例を用いて説明する．

日本文：大勢の前で話すときは多くの人があがるものだ。

日本語パターン：VP1ときはNP2がV3ものだ。

対訳英語パターン：NP2 V3 when N3 have to VP1.

対訳英語文：Many people have stage fright when they have to give a speech in front of a large audience.

下線部が，本研究の対象である動詞句である．

1章のはじめに述べたように，局所翻訳とは，英語のパターンにおいて変数で表された線形要素に，変数に対する日本語の英訳を代入することで全体の英文を生成する翻訳方式である．

動詞句についても，日本語パターンの動詞句部分に適合した日本文を用いて翻訳を行い，変数「VP」に代入する英訳を生成する．しかし，先ほどの例において，動詞句と見なされた「大勢の前で話す」という情報のみで本当に「give a speech in front of a large audience」という対訳が生成可能であるかという局所翻訳の可能性はまだ分かっていない．そこで，本研究では，その可能性の程度の検証を行う．

また，先行研究 [2] により単文の翻訳に「結合価パターン」が有効であることが分かっている．動詞句は単文とよく似ており，単文の主語を除いたものを動詞句と見なすことができる．そこで，本研究では動詞句の翻訳においても結合価パターンは有効であるかを検証する．

よって，本研究の目的は以下の2点の検証である．

- 局所翻訳の可能性の検証
- 結合価パターンの有効性検証

ここで，参考1は，結合価パターンを用いた翻訳システムである．参考1は，非線形性のある格要素と述語の関係は結合価パターンでカバーしており，さらに他の手法を用い，

線形要素についてもカバーしている．本研究では，結合価パターンのみで動詞句の翻訳がどの程度可能であるかを検証するため，本研究にて翻訳プロトタイプ「queen」の作成を行う．

3 結合価パターンを用いた動詞句翻訳

本研究では、動詞と格要素だけで動詞句の翻訳がどの程度できるかを検証するために翻訳プロトタイプ「queen」を作成した。queen は [3] を基にしている。

3.1 翻訳用知識ベース（結合価パターンの変形）

3.1.1 結合価パターンの句形成

結合価パターンは、それぞれ一文で表される。本研究では、文ではなく動詞句を扱うため、文中の主語に該当する部分の削除を行う。基本となる結合価パターンを変形した例を以下に示す。

（変更前）: $N1$ use $N2$ as $N3$

（変更後）: use $N2$ as $N3$

例のように、通常、パターンは名詞変数 N から始まり（14819 パターン中 14694 パターン）、それが主語になっているものが多い。 N で始まっていないもの（残り 125 パターン）でも、規則があるものがある。そこで、主語を除くアルゴリズムを以下のように定めた。

- 名詞変数 N から始まるもののうち、 N が所有格でないもの
名詞変数部分のみを消去する。

例)

元パターン : $N1$ become $N2$

主語を除いたパターン : become $N2$

- 名詞変数 N から始まるもののうち、 N が所有格のもの
所有格の後ろには必ず名詞がきているはずなので、名詞変数とその次の変数を消去する。

例)

元パターン : $N1$'s eyes meet $N2$

主語を除いたパターン : meet $N2$

- a, または an から始まるもの (7 パターン)

a や an の後ろには必ず名詞がきている。なので、a または an とその次の変数を消去する。

例)

元パターン：a fire break out in $N1$

主語を除いたパターン：break out in $N1$

- people で始まるもの (6 パターン)

6 パターンを見てみると、全て、まず最初に現れる名詞変数の後に主動詞がきていた。そこで、最初に現れる名詞変数以前の単語を全て消去する。

例)

元パターン：people in $N1$ be coldhearted

主語を除いたパターン：be coldhearted

- it, または there, または we で始まるパターン (78 パターン)

この3つはこれら自身が主語になっていて、次に動詞がきている。なので、これら自身を消去する。

例)

元パターン：it be a long time since $N1$

主語を除いたパターン：be a long time since $N1$

これにより、99%以上 (14785 パターン)、うまく主語を除くことができた。また、今後、過去形、三人称単数の“s”などに対応できるように、主動詞に形態素調整用のマークを付与する。主語を除いた後、パターンの先頭にマークを付けるようにしてあるので、14785 パターンを除く残り 34 パターンだけ、マーク付けがうまくいってないことになる。

3.2 補語の削除

結合価パターンと照合を行う際、結合価パターンにおいて、入力動詞句にはない補語が含まれることがある。この補語を残しておいてしまうと、出力結果に変数が字面で残ってしまう。そこで、不要な補語を削除する。例を以下に示す。

例)

入力動詞句：授業料を払う。

理想解：pay the tuition

適合パターン (日; 英): $N1$ (3 主体) が $N2$ (1168 制度 (経済) 934 貨幣) を $N3$ (3 主体) に払う。; $N1$ pay $N2$ to $N3$

代入値： $N2$ =授業料 tuition

動詞句「授業料を払う。」に対し、「 $N1(3 \text{ 主体})$ が $N2(1168 \text{ 制度 (経済) } 934 \text{ 貨幣})$ を $N3(3 \text{ 主体})$ に払う。」というパターンが適合した。用言「払う」と、更に格要素の中の“を格”である $N2$ と授業料が適合している。しかし、日本語パターンには「 $N3(3 \text{ 主体})$ に」という格要素も含まれる。これは補語であり、入力動詞句には存在しない要素である。そのため、 $N3$ は代入値を持たず、変数のまま残ってしまう。変数のまま残ってしまうと、出力が「pay tuition to $N3$ 」となってしまうが、入力動詞句に存在しない要素「 $N3(3 \text{ 主体})$ に」に対応する「to $N3$ 」を削除すれば、出力が「pay tuition」となり、入力動詞句に則した結果となることが分かる。

ここで、入力にない要素が削除の対象となるが、目的語が入力にない要素だとしても削除の対象にならない。それは、目的語は英語文において、文の骨格を成すものであるためである。入力にない要素が目的語であるか否かを判定するのは、以下のアルゴリズムを用いる。

- パターンにおいて最後に適合した変数以降に変数（所有格ではない）がそのまま残っている場合

この場合、入力動詞句は、その残った任意格には適合していないと考えられ、消しても問題はない。なので、最後に適合した変数より後に適合することなく残った変数があるなら、そのあとの文字列は消去する。例を以下に示す。

入力動詞句：それを認める。

理想解：accept that

適合パターン（日；英）： $N1(3 \text{ 主体})$ が $N2(*)$ を $N3(*)$ と認める。； $N1$ recognize $N2$ as $N3$

出力結果：@recognize it

代入値： $N2$ =それ it

例において、最後に適合した変数は $N2$ である。しかし、適合した英語のパターンには $N2$ のあとに更に $N3$ という変数がある。 $N3$ に代入値がないということは、つまり、日本語のパターンの中の「 $N3(*)$ と」の部分が入力動詞句にないことになる。よって、その対訳である「as $N3$ 」を削除する。

- パターンにおいて最後に適合した変数以降に変数が残っていない場合、または、残った変数が所有格の場合

この場合、入力動詞句は、パターンの全ての任意格に適合していると考えられる。なので、そのあとの文字列を消すことはしない。

例 1)

入力動詞句：たくさんの人に迷惑を掛ける。

理想解：annoy many people along the way

適合パターン(日;英): $N1(3 \text{ 主体})$ が $N2(3 \text{ 主体})$ に 迷惑を 掛ける。; $N1 \text{ cause } N2 \text{ trouble}$

出力結果：@cause たくさんの人 trouble

代入値： $N2$ =たくさんの人 たくさんの人

例 1 において、最後に適合した変数は $N2$ である。 $N2$ 以降、変数は存在しないため、 $N2$ 以降の文字列「trouble」は重要な情報であると考えられる。そのため、この情報を削除することはしない。

例 2)

入力動詞句：最近の報道を気に掛ける。

理想解：concern about the latest news

適合パターン(日;英): $N1(4 \text{ 人})$ が $N2(*)$ を 気に掛ける。 $N1 \text{ have } N2 \text{ on } N1's \text{ mind}$

出力結果：@have 最近の報道 on ($N1's$) mind

代入値： $N2$ =最近の報道 最近の報道

例 2 において、最後に適合した変数は $N2$ である。 $N2$ 以降に、「 $N1's$ 」という所有格が残っている。この所有格はその次の単語を修飾しており、重要な情報であると考えられる。よって、この情報を削除することはしない。

3.3 翻訳手順

翻訳アルゴリズムを例とともに以下に示す。

1. 日本語の動詞句の入力

入力する動詞句の例を以下に示す。

- ・授業料を払う。

2. 結合価パターンの抽出

入力した動詞句から、結合価パターンの抽出を行う時の条件を以下に示す。

- (a) 格要素の省略，格要素の順序変更，修飾語句の挿入，任意追加の格要素を許可する。

(b) 必修の述語付属語の条件を必須とする。

抽出したパターンの候補を以下に示す。

・ $N1$ (3 主体) が $N2$ (1168 制度 (経済) 934 貨幣) を $N3$ (3 主体) に払う。: $N1$ pay $N2$ to $N3$

・ $N1$ (4 人) が $N2$ (533 具体物 1235 事) を払う。: $N1$ get rid of $N2$

・ $N1$ (4 主体) が $N2$ (388 場所 865 家屋 (本体) 1595 宿泊) を払う。: $N1$ move to $N2$

3. 照合による選択

意味属性制約を行い，更に慣用句表現を優先的に選択し，最も広範囲に入力文と一致するパターンの選択を行う。候補のパターンから，照合により選択されたパターンを以下に示す。

・ $N1$ (4 主体) が $N2$ (1168 制度 (経済) 934 貨幣) を $N3$ (3 主体) に払う。: $N1$ pay $N2$ to $N3$

4. 名詞変数への訳語の挿入，入力にない要素の削除

既存の日英辞書引きプログラムを用い，入力動詞句と適合する変数に訳語を代入する。入力にない要素については，目的語ならばパターンの骨格を左右するので残すが，補語ならば骨格とは関係ないため削除する。名詞変数への挿入と，入力にない要素の削除を以下に示す。

・ $N2$ =授業料 tuition

また，「to $N3$ 」については，入力動詞句にない要素であり，補語であるため消去する。

5. 英語の動詞句の形成

主語に該当する変数を削除し，句を形成する。句を形成した例を以下に示す。

・ pay tuition

6. 句形成した英文の出力

句形成した英文をそのまま出力する。

このようにして，日本語の動詞句の翻訳を行う。

なお，本プロトタイプでは，主動詞と格要素の訳出構造をとらえることが目的である。名詞訳語の選択，任意の格要素（時間，場所，手段など）の訳語挿入，副詞の挿入，冠詞・数の問題は扱わない。

4 実験

4.1 実験の目的

動詞句の局所翻訳の可能性および動詞句の翻訳における結合価パターンの有効性の検証を行うことを目的とする。そのため、queen、参考1、参考2による動詞句のそれぞれの訳出を出力する。そして、訳出品質の評価値を付与する。

4.2 翻訳対象

[1]では、重文・複文の15万文対の日英対訳コーパスから文型パターンを作成した。その作成過程において、対応関係の見出された動詞句が約7万件存在する。

本研究では、この対訳コーパスから、ランダムに取り出した325句対の動詞句を翻訳対象とする。ただし、英語句を構成する単語数ごとに分類し、その中で2単語から14単語までのものを、それぞれランダムで25句ずつ取り出した。

4.3 訳文の評価

4.3.1 訳文の評価値と評価基準

本研究の評価値と評価基準を以下に示す。

評価⁺：主動詞と主名詞が理想解と完全一致。

評価⁺：理想解と異なる構文だが、意味は合っている。

評価⁺：基本的な構造は良いが、得たパターン、または得た英文の字面の一部を変形する必要がある。

評価^x：訳出が間違っている。

評価⁻：出力パターンが一つもない。

ただし、線形翻訳で追加できる前置詞句は問わない。

以上の計5段階にて人手により評価を行う。また、queenにおいて、結果が複数個出力された場合は、全ての結果に評価値を付与した後、最も最良の評価値をその出力結果の評価値とする。

また、5段階の評価において、正解と見なすものは、評価⁺ または評価⁺ である。

4.3.2 評価方法

評価方法として、再現率・適合率を用いる。定義を以下に示す。

再現率 R : 全入力事例数における、出力パターンが一つでもある事例数 N の割合

適合率 P : N における、評価値が または である事例数 (スコア) の割合

4.3.3 評価の例

queen の評価の例を以下に示す。

- 評価 の例

入力動詞句：パソコンを使う。

理想解：use personal computer

適合パターン (日 ; 英) : $N1$ (3 主体 760 人工物 *) が $N2$ (*) を $N3$ (*) に 使う。 ;

$N1$ use $N2$ as $N3$

出力結果：@use personal computer

代入値： $N2$ =パソコン personal computer

理想解と、主動詞 (use) と主名詞 (personal computer) が完全に一致しているので評価 にする。

- 評価 の例

入力動詞句：病気に成る。

理想解：result in her illness

適合パターン (日 ; 英) : $N1$ (4 人 536 動物 (個体)) が病気に 成る。 ; $N1$ fall sick

出力結果：@fall sick

理想解とは異なる構文であるが、入力動詞句に対して出力の意味は通っているので評価 にする。

- 評価 の例

入力動詞句：都会に移る。

理想解：move to the city

適合パターン (日 ; 英) : $N1$ (3 主体 389 施設 533 具体物) が $N2$ (458 地域 468 自然 533 具体物 2610 場) から/より/を $N3$ (458 地域 468 自然 533 具体物 2610 場) に/へ 移る。 ; $N1$ move from $N2$ to $N3$

出力結果：@move from ($N2$) to city

代入値：N3=都会 city

基本的な構造は良いが，出力において変更しなければならない部分がある．なお，下線部が変更しなければならない字面の部分であるので評価にする．

- 評価×の例

入力動詞句：足に怪我をする。

理想解：hurt my leg

適合パターン(日；英)：N1(3主体)がN2(*)をN3(*)に/とする。；N1 make N2 N3

出力結果：@make accident foot

代入値：N3=足 foot，N2=怪我 accident

入力動詞句に対し，出力結果が正しくないので×にする．

- 評価-の例

入力動詞句：頭から追い出す。

理想解：get out of our heads

適合パターン：なし

入力動詞句に対し，出力が一つも存在しないので-にする．

4.4 実験結果

queen，参考1，参考2のそれぞれの評価の付与数の結果を表2に示す．下段は，出力があったものの中のそれぞれの評価の付与数の割合である．

表 2: 3つのシステムの翻訳結果

	評価の付与数 (出力結果あり中の評価の付与数の割合)				
				×	-
queen	86 (29.6%)	135 (46.4%)	36 (12.4%)	34 (11.7%)	34
参考1	74 (22.8%)	156 (48%)	24 (7.4%)	69 (21.2%)	0
参考2	84 (25.8%)	162 (49.8%)	25 (7.7%)	54 (16.6%)	0

queen と、他の 2 つのシステムを比べると、queen は評価 の割合が多いことが分かった。評価 とは、「基本的な構造は良いが、得たパターンの字面の一部を変形する必要がある」というものなので、字面を変形することがうまくできれば、正解と見なすことができるようになる。例を以下に示す。

入力動詞句：都会に移る。

理想解：move to the city

適合パターン (日; 英): $N1(3 \text{ 主体 } 389 \text{ 施設 } 533 \text{ 具体物})$ が $N2(458 \text{ 地域 } 468 \text{ 自然 } 533 \text{ 具体物 } 2610 \text{ 場})$ から/より/を $N3(458 \text{ 地域 } 468 \text{ 自然 } 533 \text{ 具体物 } 2610 \text{ 場})$ に/へ 移る。;
 $N1 \text{ move from } N2 \text{ to } N3$

出力結果：@move from ($N2$) to city

代入値： $N3$ =都会 city

基本的な構造は良いが、出力において変更しなければならない部分がある。なお、下線部が変更しなければならない字面の部分である。ここがうまく変形することができれば、評価 や になる。

queen, 参考 1, 参考 2 の再現率 R と適合率 P の結果を表 3 に示す。

表 3: queen, 参考 1, 参考 2 の再現率 R と適合率 P

	queen	参考 1	参考 2
再現率 R	89.5%(291/325)	100%(325/325)	100%(325/325)
適合率 P	75.9%(221/291)	70.1%(230/325)	75.7%(246/325)

queen は、結合価パターン辞書のみを使用して翻訳を行っている。結合価パターンにない動詞の追加、意味属性制約の柔軟な対処などを行い、また、結合価パターン辞書の範囲外である、副詞についても実装し、格要素内の名詞を名詞句として認識できるようにするなど、更に改良を加えることにより、再現率、適合率ともに向上すると思われる。

5 考察

5.1 局所翻訳の可能性の検証について

局所翻訳の可能性については、4.4 節の結果で示した 3 つのシステムの再現率 R 、適合率 P にて評価ができると思われる。よって、現段階で、約 70% ~ 76% であることが分かった。また、入力動詞句に対する英語の理想解の構成単語数ごとのスコア (0 から 1 に正規化したもの) を図 2 に示す。

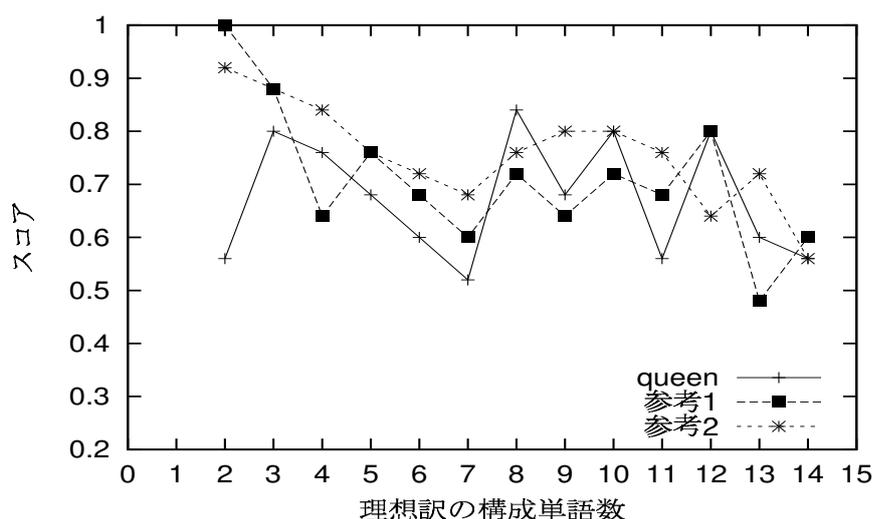


図 2: 各単語数でのスコア

図 2 より、構成単語数が 2 単語から 7 単語に近づくにつれて、スコアが減少する傾向がある。しかし、7 単語を過ぎると、0.5 から 0.8 までの範囲に収まっている。

これは、主動詞と主名詞の構造を中心に評価したためと考えられる。今後の課題に、その他の部分の訳出構造を検討することがあげられる。

5.2 結合価パターンの有効性の検証について

queen の翻訳実験において、評価が または となった正解となる事例数は、全体の 68% (221/325) であった。よって、現段階では結合価パターンは 68% 有効性があると言える。

本研究は、翻訳プロトタイプの第一段階であり、4.4 節の結果で述べたように、今後さ

らに改良の余地があると思われる。そのため、第一段階としては、良い結果だったと言える。

以下では、評価が正解とならなかった計 104 件を分析する。

分析の結果、動詞句の翻訳においては、結合価パターンで実装されている格要素と動詞の関係だけでは判断できない非線形と思われるものには、以下のものがあった。

1. 格要素がなく、「副詞 + 動詞」で表されている場合
2. 格要素が足りない場合
3. 1 でも 2 でもなく、動詞だけで適合した場合
4. 格要素が複雑である場合
5. 入力動詞句が「A を B にする」構文である場合
6. 消去すると意味を失う構文
7. 変数に誤った単語が代入された場合
8. 再帰代名詞の変数が残っている場合
9. 主語となる変数に適合した場合
10. 結合価パターンに存在しない場合
11. 意味属性制限ミス

その具体例を以下に示す。

1. 格要素がなく、「副詞 + 動詞」で表されている場合
格要素がないので、うまく訳出できない。

例 1)

入力動詞句：うまく行く。

理想解：go well

2. 格要素が足りない場合

パターンには存在する格要素となる変数が入力動詞句にない場合、変数がそのまま残ってしまう。

例)

入力動詞句：北に変わる。

理想解：shift to the north

適合パターン(日;英): $N1(*)$ が $N2(*)$ から $N3$ に $(*)$ 変わる。; $N1$ change from $N2$ to/into $N3$

出力結果：@change from ($N2$) to/into north

代入値： $N3$ =北 north

また、この場合において、足りない格要素が目的語であった場合

例)

入力動詞句：型紙に合わせる。

理想解：fit the pattern

適合パターン(日;英): $N1(3$ 主体) が $N2(533$ 具体物 2306 物象) を $N3(2306$ 物象 2488 性格 533 具体物) に/と合わせる。; $N1$ match $N2$ with $N3$

出力結果：@match ($N2$) with pattern

代入値： $N3$ =型紙 pattern

英語の動詞が他動詞であるにも関わらず、入力文に該当格要素が存在しなかった。つまり、「合わせる」の目的語の情報がない。これは、英文生成時の問題とすることもできる。

3. 1でも2でもなく、動詞だけで適合した場合

格要素は関係なく、動詞だけ結合価パターンと適合してしまったものがあった。

例)

入力動詞句：ダム構造と水の重量に左右する。

理想解：depend on the weight of the structure and of the water on the dam

適合パターン(日;英): $N1(*)$ が $N2(*)$ を左右する。; $N1$ influence $N2$

出力結果：@influence ($N2$)

「ダム構造と水の重量に」の部分が全て離散記号と適合してしまった。

4. 格要素が複雑である場合

例)

入力動詞句：異なる字体の文字やグラフィックスまでも出力できる。

理想解：are capable of producing characters in different type fonts and even graphic output

この場合は、格要素が複雑であるため、「*A* を出力できる = are capable of producing *A*」というパターンがあったとしても、訳出は難しい。

5. 入力動詞句が「*A* を *B* にする」構文である場合

「*A* を *B* にする」は、「make *A B*」という対訳の結合価パターンを持つ。入力動詞句が「*A* を *B* にする」構文の時、3.3 節の翻訳手順の「照合による選択」により、全て「make *A B*」という出力結果になってしまう。

例)

入力動詞句：足に怪我をする。

理想解：hurt my leg

適合パターン(日;英): $N1(3 \text{ 主体})$ が $N2(*)$ を $N3(*)$ に/と する。; $N1$ make $N2$ $N3$

出力結果：@make accident foot

代入値： $N2$ =怪我 accident, $N3$ =足 foot

6. 消去すると意味を失う構文

本稿では、3.2 節の補語の削除において、入力動詞句と適合しない残った変数が補語であれば、消去している。しかし、消去することによって本来の意味を失う構文がある。

例)

入力動詞句：日よけを上げる。

理想解：put up their sunshade

適合パターン(日;英): $N1(4 \text{ 人})$ が $N2(533 \text{ 具体物 } 3 \text{ 主体})$ を $N3(“ \text{ リスト/上 ” } 875 \text{ 屋根})$ に 上げる。; $N1$ put $N2$ on $N3$

出力結果：@put sunshade

代入値： $N2$ =日よけ sunshade

この例は、「put *A* on *B*」で初めて「上げる」という意味になる。今後、何とか動詞の判定を行う必要がある。

7. 変数に誤った単語が代入された場合

queen が格要素の字面を誤って認識していて、変数に誤った単語が代入されてしまう。

例)

入力動詞句：写真をテーブルの上に並べる。

理想解：line the photos out on the table

適合パターン（日；英）：N1(3 主体 962 機械) が N2(4 人 389 施設 533 具体物) を N3(388 場所 533 具体物 2610 場) に/へ 並べる。； N1 arrange N2 on/in N3

出力結果：@arrange photograph on/in head

代入値：N2=写真 photograph, N3=上 head

本来, N3 には「テーブル」があたるべきである。

8. 再帰代名詞の変数が残っている場合

入力動詞句だけでは判断不可能で, 変数が残ってしまう。

例)

入力動詞句：人間に化ける。

理想解：take the shape of a human being

適合パターン（日；英）：N1(533 具体物) が N2(533 具体物) に 化ける。； N1 disguise N1-self as N2

出力結果：@disguise (N1)-self as human being

代入値：N2=人間 human being

この場合は, 格要素と述語の関係だけで決定することができない。また, この問題は, 局所翻訳の可能性の問題でもある。つまり, N1-self の訳出は, 主語の情報が必要である。動詞句の翻訳でも, 主語を入力する必要がある。

9. 主語となる変数に適合した場合

queen では, 3.1.1 節の結合価パターンの句形成において, 主語となる変数は削除している。よって, 主語となる変数に適合した情報は, 削除されてしまう。

例)

入力動詞句：他の鑄鉄とは含有グラファイトの形状が異なる。

理想解：differ from other cast irons in the shape of the contained graphite

適合パターン（日；英）：N1(*) が N2(*) と 異なる。； N1 be different from N2

出力結果：@be different from cast iron

代入値：N1=形状 shape, N2=鑄鉄 cast iron

主語に該当する変数に日本語動詞句の一部が適合しているが, 主語は消去するため, 出力では主語の部分の情報が失われる。

10. 結合価パターンに存在しない場合

結合価パターンに存在しないため, うまく適合しない。内訳は以下の通り。

主動詞が「名詞＋する」である場合，主動詞が複合動詞である場合，結合価パターンにない動詞の場合

11. 意味属性制限ミス

内訳は以下の通り．

意味属性制限がうまくいかない場合，意味属性制限がかかり過ぎて適合するパターンがない場合

参考として，動詞と格要素の関係において，判別できるはずのものは，104 件中 73 件であった．queen のパターンの照合条件の修正，結合価パターンの追加，意味属性制約への柔軟な対処が求められる．これが実装できれば，結合価パターンの有効性が約 90%まで上がると考えられる．

また，結合価パターンで表現できない非線形構造については，別途パターン化する必要が考えられる．

5.3 諸般の問題

5.3.1 補語の削除の問題

5.2 節の結合価パターンの有効性の検証の「6. 消去すると意味を失う構文」において，消去することによりパターンの本来の意味を失う構文があることが分かった．今後，何とか動詞の判定を行う必要がある．

5.3.2 対訳英語が 1 単語で構成される場合

本研究では，実験対象として，理想解の英語句の構成単語数が 2 単語から 14 単語のものを対象とした．つまり，1 単語は対象外である．文型パターンから日英の対訳の見出された約 7 万件の動詞句のうち，1 単語で構成されるものは 4374 件あった．1 単語を対象外とした理由は，1 単語は句ではなく単語であるためである．しかし，テストデータとして，1 単語からなるもの 20 件を用いて実験を行った．それについて報告する．20 件での翻訳実験の結果を表 4 に示す．

表 4: 理想解の構成単語が 1 の時の実験結果

評価の付与数				
			×	-
0	7	9	3	1

いずれの結果についても，出力結果が 1 単語で構成されるものはなかった．20 件について，気づいたことを以下に示す．

入力動詞句について，「一つの格要素 + 動詞」または「副詞 + 動詞」のいずれかであった．この 3 つについて，具体的に述べる．

- 「一つの格要素 + 動詞」の場合 (15 件)

適合したパターンの名詞変数と格要素が適合している場合は，うまく翻訳できていた．例を以下に示す．

例)

入力動詞句：手紙を書く。

理想解：write

適合パターン(日；英)： $N1(3 \text{ 主体})$ が $N2(*)$ を $N3(1109 \text{ 文書 } 1064 \text{ 言葉})$ に/で書く。； $N1 \text{ write } N2 \text{ in } N3$

出力結果：@write letter

代入値： $N2$ =手紙 letter

入力動詞句において，一つの格要素となっている「手紙を」とパターンの $N2$ が適合している．不要な補語「in $N3$ 」を削除して，出力は意味の通る結果となっている．

一つの格要素がパターンにおいて慣用表現になっており，パターン内にある名詞変数が全て残ってしまう場合は，うまく翻訳できていない．例を以下に示す．

例)

入力動詞句：鼻を鳴らす。

理想解：grunt

適合パターン(日；英)： $N1(4 \text{ 人 } 535 \text{ 動物})$ が $N2(4 \text{ 人 } 535 \text{ 動物})$ に鼻を鳴らす。； $N1 \text{ whine to } N2$

出力結果：@whine to ($N2$)

入力動詞句の全てが、パターンの慣用表現に適合している。よって、パターンに含まれる名詞変数に適合する日本語がない。名詞変数が目的語か補語か判断できないため、削除することもできないため、そのまま変数が字面で残ってしまう。

- 「副詞 + 動詞」の場合（5件）

日本語表現が「副詞 + 動詞」で表すものは、副詞の部分が訳出されない。例を以下に示す。

例)

入力動詞句：ガタガタと震える。

理想解：shudder

適合パターン（日；英）： $N1$ (4人 535 動物) は $N2$ (552 動物 (部分) が $N3$ (声) から/より/で 震える。； $N1$'s $N2$ tremble with/from $N3$

出力結果：@tremble with/from ($N3$)

結合価パターンは、用言と格要素の意味的關係を記述したものであるため、入力動詞句に副詞が含まれても、その情報が翻訳されることはない。

6 おわりに

5.1 節の局所翻訳の可能性の検証と、5.2 節の結合価パターンの有効性の検証の結果より、主動詞と主名詞の構造に関して、文全体から見て線形要素とする動詞句については、代入された日本語だけからの翻訳の可能性が約 70% ~ 76% であることが分かった。また、動詞句の翻訳においても、文における主語などの情報が必要な時があることも分かった。そして、結合価パターンの有効性は、現段階で 68% であることが分かった。queen のパターンの照合条件の修正、結合価パターンに存在しない動詞の追加、意味属性制約への柔軟な対処を行うことにより、結合価パターンの有効性が約 90% まで上がると考えられる。また、格要素と動詞の共起から生じるものの他に、副詞との共起、格要素自体の複雑さが問題になることが分かった。今後、より詳細に分析を進める。

謝辞

最後に，本研究を遂行するにあたり，御指導いただきました鳥取大学工学部知能情報工学科計算機 C 研究室の池原教授，村上助教授，徳久助手に厚くお礼申し上げます．

また本研究で用いた「ALT-J/E」は NTT との共同研究の関係から使用させていただきました．ここにお礼申し上げます．

参考文献

- [1] 池原悟, 阿部さつき, 徳久雅人, 村上仁一: 非線形な表現構造に着目した重文と復文の日英文型パターン化, 自然言語処理, Vol.11, No.3, pp69-95, 2004.
- [2] 金出地真人, 徳久雅人, 池原悟, 村上仁一: 結合価文法による動詞の訳語選択能力の評価, 自然言語処理, Vol.11, No.3, pp149-164, 2004.
- [3] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦: 日本語語彙大系, 岩波書店, 1997.
- [4] 徳久雅人, 守谷有司, 村上仁一, 池原悟: 意味属性の共起による「AのB」型名詞句の翻訳規則, 情報科学技術フォーラム (FIT-2003) 情報レターズ, Vol.2, pp.87-88.
- [5] 石上真理子, 徳久雅人, 村上仁一, 池原悟: 結合価パターンを用いた動詞句の翻訳可能性の調査, 言語処理学会第11回年次大会, 2005(発表予定).