概要

近年，機械翻訳では，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を用いる方式が期待されている．この翻訳は，日本語の入力文に合った日本語パターンに対応する英語パターンを用いて出力英文を作成する方式である．現在，重文・複文において文型パターン辞書が構築されている．この辞書では，パターン記述要素とし，名詞句変数が使われており，その変数に代入された日本語表現を英訳する必要がある．

そこで，本研究では，名詞句に対しても「文型パターン」を用いた翻訳の実現を試み，それに必要な「名詞句パターン辞書」の構築を目的とする．そして，作成した名詞句パターン辞書を使い翻訳実験することで，その精度を調べる．具体的な翻訳内容は、名詞句対訳コーパスよりランダムで日本語名詞句を 100 件選び，翻訳プロトタイプシステム (Meijin) により翻訳実験を行った．また，翻訳精度のの比較をするため，既存の翻訳機,「ALT-J/E」,「翻訳の王様」についても，Meijin と同じ入力名詞句を入力し訳出した．

評価は，再現率 R と適合率 P を用いて集計した．結果は，翻訳プロトタイプシステム「Meijin」の再現率が 75%，適合率が 99%,「ALT-J/E」の再現率が 100%，適合率が 87%,「翻訳の王様」の再現率が 100%，適合率が 94%であった．適合率において，高い数値が出ていた．これは，パターンの候補があった名詞句においての正解の割合であり，パターンの候補があった名詞句に関しては，質の良い翻訳ができたと言える．よって，名詞句パターン辞書のパターン対が良質であったとが確認できた．

目 次

図 目 次

表 目 次

1　はじめに

機械翻訳では，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を用いる方式がある．この方式では，大規模な文型パターン辞書の開発は困難と考えられ，対象分野を限定した用例翻訳などが試みられてきた．しかし，近年，大量のパターンを作ることが可能となり，最近では，重文・複文の日英機械翻訳にこの「文型パターン」を用いる手法が提案されている [1]．その手法では，パターンの記述要素に名詞句変数が使われており，その変数に代入された日本語表現を英訳する必要がある．

そこで，本研究では，名詞句翻訳においてもパターン翻訳による実現に向けて「名詞句パターン辞書」の構築を行うことを目的とする．具体的に，まず，名詞句対訳コーパスにおける全ての名詞句対に対し，単語アライメントをとることにより，名詞句パターンを作成する．それをまとめて，名詞句パターン辞書とする．そして，実際にこの辞書を用い，翻訳実験を行うことで「名詞句パターン辞書」の精度を評価する．翻訳実験は，翻訳プロトタイプシステム「Meijin」により行う．以下にその流れについて示す．まず，入力日本語名詞句と，日本語パターンを照合し，適合した日本語パターンを抽出する．次に，作成した「名詞句パターン辞書」から抽出した日本語パターンに対応する英語パターンを検索する．検索結果より，抽出された英語パターンの変数部へ単語を代入し，英語名詞句を出力する．この結果を人手で評価する．

本論文の構成は次のようになる．2 章で，研究の背景について説明する．3 章で名詞句パターン辞書の作成を説明し，4 章で翻訳手順について，5 章で翻訳実験について説明する．6 章では実験の考察を述べ，7 章では，結論および今後の課題を述べる．

2　研究の背景

2.1　等価的類推思考による機械翻訳

要素合成法を基本とする従来の自然言語処理の限界を克服することを狙いとした１つの原理に,「等価的類推思考の原理」がある．この原理に基づく翻訳方式では，翻訳対象となる両言語で言語表現を「文型パターン」によって記述しておき，意味的に等価な文型パターンを対応付けることで，意味の失われない解析・生成を実現しようとしている．

近年まで，大規模な文型パターン辞書の開発は困難と考えられ，対象分野を限定した用例翻訳などが試みられてきた．しかし，標本文には多くの線形要素が存在すること，文型パターン化作業の大半が機械化できることが分かり，大量のパターンを作ることが可能となった．

2.2　名詞句対訳コーパス

最近，重文・複文において,「文型パターン」を用いた手法が提案されている [1]．この手法により，現在，重文複文において約 15 万文の対訳データが存在する．この中で名詞句変数が使われており，その部分を抽出すると，日本語名詞句とそれに対応する英語名詞句が 45,676 対あった．この名詞句対をまとめて，名詞句対訳コーパスとする．

表 1: 名詞句対訳コーパスの一部

日本語	英語
彼のお母さん	his mother
あの建物	that bilding
あの男	that man
古い民謡の一つ	one of the old folk songs
息子の話	son's story
その画家	that painter
唯一の動物	the only animal

2.3　言語解析ツール

本研究では，既存のツール「ALT-JAWS」,「tokepie」,「Brill」,「Bickey」を使用する．以下にそれぞれの説明をする．

ALT-JAWS

NTT による日本語の形態素解析を行うツールである．与えられた文字列を単語に分解し，各単語の品詞を出力する．

tokepie

本研究室で作成した英語表現を形態素解析するツールである．タグを付与し，格変化している動詞や名詞の原形を出力できる．

Brill

参考文献 [2] の英語の形態素解析をするツールである．本研究では，tokepie より付与された複数の品詞タグを一意に絞る．

Bickey

本研究室で作成した辞書引きツールである．単語アライメントをとる際，この辞書を使用する．この辞書は，既存の電子辞書を用いて作成されている．約 12 万語を収録している．

2.4 本研究の目的

約 4.5 万件の名詞句対訳コーパスにおける，全ての名詞句対に対し，単語アライメントをとることによって，自動的にパターンを作成する．そして，それをまとめて,「名詞句パターン辞書」として構築することが目的である．

3　名詞句パターン辞書の作成

3.1　名詞句パターン化の方法

3.1.1　パターン化の手順

パターン化には，単語アライメントによる対応要素の変数化，変数への意味属性制約の付与，形態素調整用タグの付与の大きく３つの手順がある．

次にパターン化の流れを示し，3.1.2 節で単語アライメントについて，3.1.3 節で単語の変数化について，3.1.4 節で意味属性の付与について，3.1.4 節で形態素調整について詳しく述べる．

元の句

日本語名詞句：彼のお母さん

英語名詞句 ：his mother

単語アライメント

彼 $\longleftrightarrow$ his　　お母さん $\longleftrightarrow$ mother

変数化

彼 $\rightarrow PRN$　　お母さん $\rightarrow N$

his $\rightarrow PRN$　　mother $\rightarrow N$

意味属性の付与と形態素調整

変数化する際，“お母さん” は一般名詞 (N) なので，意味属性 “(男女，母)“を付与する．また，“his” は，“he” の所有格となっているので，“his” の変数 “PRN” に “^poss” を付与する．

パターン化

日本語パターン：$PRN1$ の $N2$(男女, 親)

英語側パターン：$PRN1$^poss $N2$

() の中は意味属性の意味を表す.

3.1.2　単語アライメント

単語アライメントとは，日本語表現，英語表現において，単語対応をとることである．

まず，日本語表現，英語表現それぞれで形態素解析を行う．次に，形態素解析結果から，辞書 “bickey” を用いて対応をとる．以下に例を示す.

（日本語表現）彼女の指
（英語表現）her fingers

この例において，まず日本語で形態素解析を行う．

－ 1. /彼女 (1710,NI:24,NI:49,KR:0601s07,KR:9901s88)

－ 2. +の (7410)

－ 3. /指 (1100,NI:598,NI:607,NI:608,KR:5201k00)

この１，３の単語において英語表現と対応をとる．英語表現においても tokepie を使い形態素解析を行う．以下に示す．

1/. her(1421)

2/. fingers(1122 finger)(2102 finger)

() 内は意味属性であり，“1421” は，“代名詞” を，“1122” は “名詞” を，“2102” は “動詞” を示している．ここで，brill を使い品詞を一意に絞ると以下のようになる．

1. her(1421)

2. fingers(1122 finger)

そして，対応辞書 bickey を用いて，“彼女”，“指” の対訳を検索する．

表 2: 対訳辞書 (bickey)：part1

日本語	対訳英語
彼女	her，she
指	digit，finger

上記の辞書より，“彼女 ↔her” の対応が，“指 ↔finger” の対応が上手くとれる．

しかし，アライメントが上手くとれない問題もある．その問題について以下に示す．

1．対応辞書の問題

対応辞書に無い単語は，アライメントがとれない．例えば，データに，日本語表現 “悪事”，英語表現で “crime” があり，変数化を行いたいが，対訳辞書では，“悪事” に対して表 3 に示す英語しかないため，対応がとれない．

表 3: 対訳辞書 (bickey)：part2

日本語	対訳英語
悪事	evil, ill, misdoing, perpetration, roguery, vice, villainy, wrong

2. 複合名詞の問題

複合名詞の場合，きちんとした対応がとれない問題がある．例を以下に示す．

（日本語表現）彼の血圧

（英語表現）his blood pressure

このアライメントをとると,「彼 ↔his」,「血圧 ↔pressure」の対応となった．つまり，本来 “血圧＝ blood pressure” の対応であるべき箇所がうまく対応できていない．これは，辞書「bickey」に血圧の対訳として “blood pressure” が存在しなかったためである．さらに，存在したとしても，先に当たった方が優先されるため，うまく対応が取れない可能性が高い．

3. 日本語表現と英語表現の問題

日本語表現と英語表現には双方さまざまな表現の仕方がある．例えば，日本語表現 “前科のある男” の場合，他の日本語表現として，“前科者”，英語表現では，“ a man with convictions”，“ex-convict” などがある．今回，日本語表現 “前科のある男” の対応として英語表現 “ex-convict” があった．まず，この日本語表現を形態素解析すると以下のようになる．

- 1. /前科
- 2. +の
- 3. /ある
- 4. /男

１（前科），３（ある），４（男）の単語において，英語表現と対応をとるが，英語表現の方を見るとの１単語 “ex-convict”，意味 “前科者” を用いており，1，3，4 共に対応がとれなかった．

3.1.3 単語の変数化

対応がとれた単語において，変数化を行う．変数化は，表 4 に従う．例えば，“花” という単語は，形態素解析の結果，一般の名詞である．これは，表 4 より，“N” と判定される．

表 4: 自立語の変数化

単語の品詞	品詞変数	
	日本語	英語
用言性名詞	NS	無し
数詞	NUM	NUM
代名詞	PRN	PRN
一般の名詞	N	N
動詞	無し	V
形容詞	AJ	AJ
形容動詞	AJV	無し
副詞	ADV	ADV
連体詞	REN	無し
数助詞	$UNIT$*	無し

（*$UNIT$ は，辞典 [4] に収録されている語を対象とする）

今回，助数詞変数 UNIT を導入した．これは，日本語表現の数助詞は，英語表現に反映されないことが多いためである．この変数に関しては，アライメントをとる段階で日本語表現にあり，英語表現になくてもよいとした．それは，入力した日本語表現において数助詞があった場合，作成したパターンにある数助詞と全く同じ数助詞がなければ，パターンの候補がなくなっていまうためである．いくつかの例を表 5 に示す．

表 5: 数助詞のつく名詞句

日本語	英語
２人の男	two men
その２台の車	the two cars
３つの異なるやり方	three different systems
４個の単結合	four single bonds
これら爆撃機の１３機	thirteen of the bombers

3.1.4 意味属性の付与

変数化の際，一般の名詞変数 N と，用言性名詞 NS に意味属性を付与する．これは，名詞には，語義が多いため，パターンとの照合を行う際に，適切ではないパターンが多く抽出される．これを防ぐために，今回意味属性を付与した．意味属性は,「日本語語彙大系」[3] で定義された約 2,700 の単語意味属性を使用する．

3.1.5 形態素調整の付与

変数化の際，所有格である英単語に対しては，“^poss” を，複数である単語に対しては，“^pl” を付与する．例えば，“his” は “he” の所有格であり，これを変数化すると，“PRN” となり，形態素調整を行うと “PRN^poss” となる．

3.2 名詞句パターン化の結果

名詞句コーパスから変数化できた句は，36,729 対， 字面の句は，8,947 対であった．後者は字面パターンとみなす．パターン化ができた名詞句に関してその一部を表 6 に，パターン化ができなかった名詞句に関してその一部を表 7 に示す．英語斜体で書かれた語は，表 4 に示した変数であり，() 内は，意味属性である．“^poss”，“^pl” は，形態素調整用タグで 2.2 節で示している．

表 6: パターン化ができた結果の一部

元の日本語名詞句	元の英語名詞句	日本語パターン	英語パターン
彼のお母さん	his mother	$PRN1$ の $N2$(女, 母)	$N1$^poss $N2$
寺の鐘	the temple bell	$N1$(団, 寺) の $N2$(文具)	the $N1$ N2
悲しい話	sad tales	$AJ1N2$(説話, 言葉, 話題等, 発言)	$AJ1$ $N2$^pl
かなり長い間	quite a long time	$ADV1AJ2N3$(関係, 中心・周辺, タイム, 時機)	$ADV1$ a $AJ2$ $N3$
その刀	the sword	その $N1$(道具)	the $N1$

表 7: パターン化ができなかった結果の一部

日本語名詞句	英語名詞句
前科のある男	an ex-convict
数多くの使い方	a multitude of uses
何百人もの人	hundreds of people
急ぎの用	urgent business
大きな資本	a lot of capital

パターン化ができなかった原因は，3.1 節で示した問題と同じである．

3.3 名詞句パターン辞書の作成結果

日英名詞句パターン対において，同じ記述のパターン対を 1 つにまとめて，パターン辞書とする．日本語パターンは，字面パターンを含め，全部で 23,834 種類あった．日本語名詞句の圧縮率は，52%であった．なお，意味属性，形態素調整を付与したままである場合は，全部で 35,289 種類で，圧縮率 22.7%であった．

パターンを作るために用いたコーパスの名詞句の分布を調べたところ，パターン化の元の名詞句が一番多く使われていたパターンが $REN1N2$ であり，その名詞句の数は 3,719 個であった．以下，上位 10 位までの日本語パターンと，それに対する英語パターンの頻度の多かった上位 2 位とその他を表 8 に示す．

コーパスの名詞句が 1,000 個以上であったパターンが 4 件，999〜100 個であったパターンが 14 件，99〜20 個であったパターンが 56 件，29〜1 個であったパターンが 23,735 件であった．この日本語パターンに対する英語パターンの頻度を図 1 に示す．

表 8 を見ると，英語パターンが第 1 位の割合が高い日本語パターンにおいては，そのまま第一位の英語パターンを適応すれば，良い翻訳ができそうにみえる．しかし，“N1 の N2” や “AJ1N2” などは，対応する英語パターンにおいてばらつきがあり，英語パターンの選択をする必要がある．

表 8: 日本語パターンに対する英語パターンの種類（日本語の上位 10 位まで）

日本語パターン 句の数	英語パターン			
	1 位	2 位	3 位	その他 [種類数]
$REN1N2$ 3,719 個	$PRN1$ $N2$ (87.0%)	$AJ1$ $N2$ (5.6%)	the $AJ1$ $N2$ (1.7%)	その他 [101] (5.7%)
その $N1$ 3,686 個	the $N1$ (97.2%)	his $N1$ (0.4%)	this $N1$ (0.2%)	その他 [41] (2.2%)
$PRN1$ の $N2$ 1,936 個	$PRN1$ $N2$ (97.2%)	$PRN1$ true $N2$ (0.2%)	$N2$ of $PRN1$ (0.1%)	その他 [38] (2.5%)
$N1$ の $N2$ 1,224 個	the $N2$ of the $N1$ (12.5%)	$N1$ $N2$ (11.8%)	the $N1$ $N2$ (10.1%)	その他 [186] (65.6%)
この $N1$ 719 個	the $N1$ (95.3%)	$N1$ (0.7%)	those $N1$ (0.4%)	その他 [17] (3.6%)
$PRN1$ の $NS2$ 661 個	$PRN1$ $N2$ (99.2%)	$PRN1$ own $N2$ (0.2%)	$N2$ of $PRN1$ (0.2%)	その他 [6] (0.4%)
$AJ1N2$ 524 個	$AJ1N2$ (46.1%)	a $AJ1$ $N2$ (34.9%)	the $AJ1$ $N2$ (8.0%)	その他 [35] (11.0%)
その $NS1$ 496 個	the $N1$ (97.3%)	their $N1$ (0.4%)	my $N1$ (0.4%)	その他 [8] (1.9%)
$REN1NS2$ 461 個	$PRN1$ $N2$ (76.4%)	$AJ1$ $N2$ (10.6%)	the $AJ1$ $N2$ (2.0%)	その他 [21] (11.0%)
$AJV1N2$ 381 個	$AJ1N2$ (45.1%)	a $AJ1$ $N2$ (25.7%)	the $AJ1$ $N2$ (12.1%)	その他 [34] (10.8%)

表 9: 表 7 に対する上位 4 位までの例

日本語パターン	元の日本語句	英語パターン 元の英語句
$REN1N2$	あの建物	$PRN1$ $N2$ The building
	近くの病院	$AJ1$ $N2$ nearby hospital
その $N1$	その会社	the $N1$ the company
	その秘密	the $AJ1$ the secret
	その手紙	his $N1$ his letter
$PRN1$ の $N2$	彼の性格	$PRN1$ $N2$ his character
	私の過去	$PRN1$ $AJ2$ my past
$N1$ の $N2$	国の将来	the $N2$ of the $N1$ the future of the country
	列車の時間	$N1$ $N2$ train time

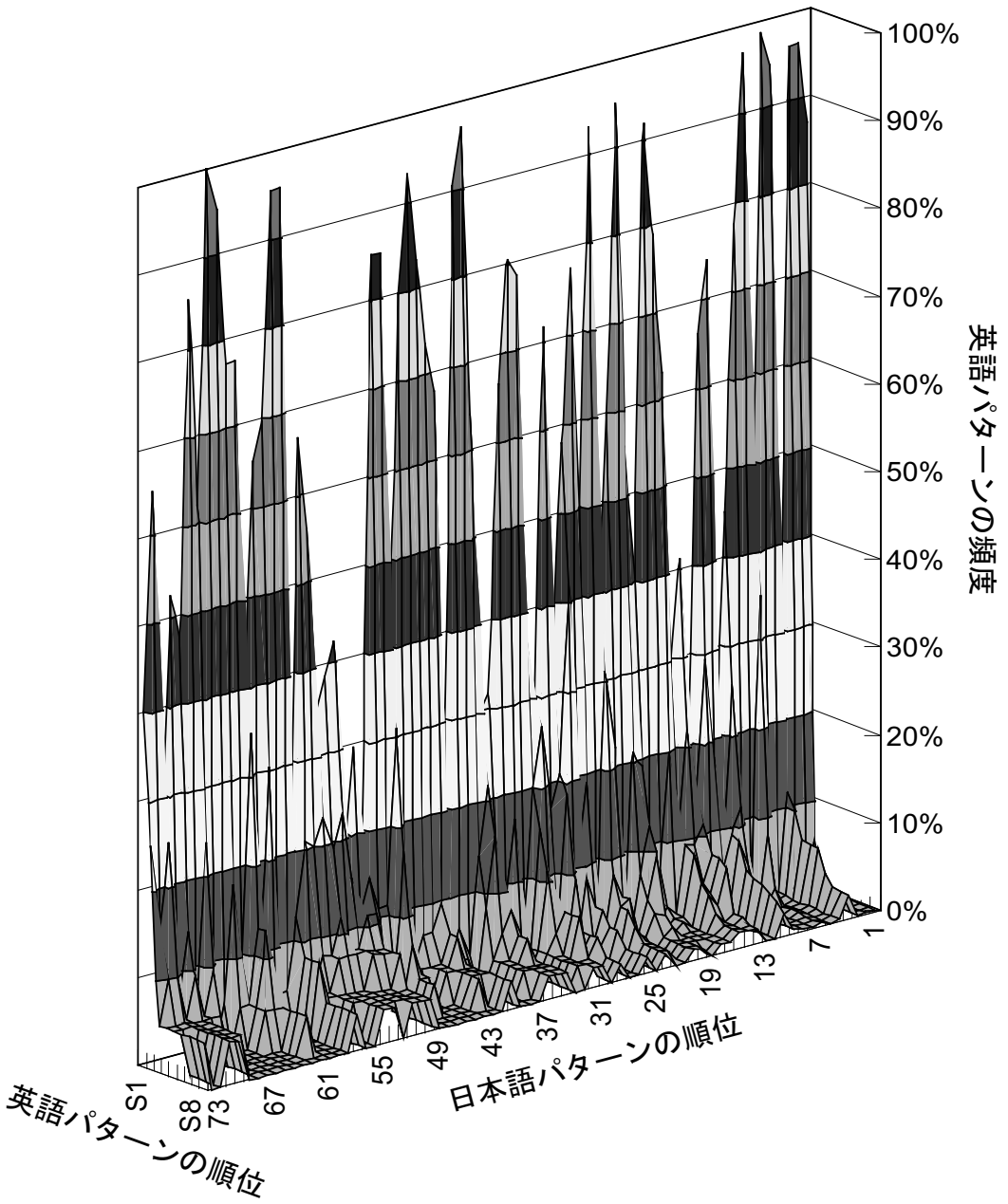

図 1: 日本語パターンに対する英語パターンの頻度

1〜73 まで書いてある軸は,「日本語パターンの順位」を表しています．1 が日本語パターン第 1 位の $REN1N2$ のことです．S1〜S8 まで書いてある軸は,「日本語パターンに対応する英語パターンの順位」で S1 が英語パターン第 1 位のことです．0〜100%まで書いてある軸は,「英語パターンの頻度」を表しています.

4　翻訳プロトタイプシステム「Meijin」による名詞句翻訳

手順を以下に示す.

1. 入力の日本語名詞句と日本語パターンをパターンパーサを用いて照合する [5].
2. 照合結果より，適合した日本語パターンを抽出する.
3. 抽出した日本語パターンに対応する英語パターンを名詞句パターン辞書から検索する.
4. 抽出された英語パターンの変数部に対応する英単語を代入し，出力する.

以下に具体例を示す.

- 入力句：この新聞
- 模範訳：this newspaper
- 日本語パターン：$REN1N2$(本)
  この $\rightarrow$ $REN1$
  新聞 $\rightarrow$ $N2$(本)
- 英語パターン：$PRN1$ $N2$
  ここで，変数部に訳語が代入される.
  $PRN1$ $\rightarrow$ this
  $N2$ $\rightarrow$ newspaper
- 出力：this newspaper

この例の場合，まず，入力した日本語句 “この新聞” と日本語パターンを照合させる．照合結果から適合した日本語パターンは，“$REN1N2$(本)” である．日本語表現 “この” と日本語パターンの連体詞変数 “REN” に，日本語表現 “新聞” が日本語パターンの一般名詞変数 “$N2$” にマッチし，さらに，() 内の意味属性 “本” に，今回 “新聞” の意味属性がマッチしたため，この日本語パターンが抽出される．この日本語パターンに対応する英語パターンを今回作成した名詞句パターン辞書を使い検索する．名詞句パターン辞書には，日本語パターンとそれに対応する英語パターンを同時に掲載しているので，検索結果より，“$REN1N2$(本)” に対応する英語パターン “$PRN1$ $N2$” が抽出される．この英語パターンにおいて，変数部に辞書引きプログラムを用い英単語を代入する．これにより，訳出された英語名詞句は，“this newspaper” である.

5　翻訳実験

5.1　実験の目的

作成した名詞句パターン辞書の性能を評価することを目的とする．具体的には，既存の翻訳機,「ALT-J/E」，および,「翻訳の王様」の翻訳精度と Meijin の翻訳精度を比較する．

5.2　実験対象

実験の入力データは，3 章で述べた日本語名詞句をランダムに選んだ 100 件を対象とする．

Meijin においては，入力された名詞句から作られるパターンは，照合に用いないこととする．4 章で述べた翻訳手順に従って訳出する．ただし，Meijin は，複数の訳出があるが，名詞句パターン辞書の性能を調べることがねらいなので，その選択は人手で行うこととする．

「ALT-J/E」,「翻訳の王様」については，Meijin と同様の名詞句を入力し訳出する．

<入力データの一部>

- この新聞
- 道路工事の騒音
- 彼の忠告
- 別の機会
- 先日の就任式
- それらの変更
- 独自の特徴

5.3 評価方法

評価基準は，以下の通りとする．また，本システム (Meijin) では，複数の訳出がでた場合，人手で最適解を決め，評価する．

表 10: 評価の定義

○	訳出された英語が，文法的に正しく，意味も理解できる場合（英語の訳語，冠詞，句の外の情報は考慮しない）
△	訳出された英語が，文法的に間違っているが，意味は理解できる場合
×	訳出された英語が，意味的に間違っている，または，訳出が無い場合

以上の評価を，再現率 R，および，適合率 P を用いて集計する．

$$\text{再現率 } R = \frac{\text{出力パターンが一つ以上ある回答の数}}{\text{出題数}}$$

$$\text{適合率 } P = \frac{\text{評価 (○) のある回答数}}{\text{出力パターンが一つ以上ある回答数}}$$

以下に評価の例を示す．

評価○の例

（入力句）別の機会

（解答例）another opportunity

（出力句）a different opportunity

（理由）訳出された句 “a different opportunity” は，文法的にも意味的にも正しいので評価○となる．

評価△の例

（入力句）新幹線の旅

（解答例）The trip by Shinkansen

（出力句）the trip of a sinkansen

（理由）訳出された句 “the trip of the Sinkansen” は，“of” が誤りであるために，評価は△となる．

評価×の例

（入力句）あの俳優

（解答例）that actor

（出力句）that sumo actor

（理由）訳出された句 “that sumo actor” は，意味が明らかに異なるので，評価×となる．

5.4 実験結果

評価結果を表 11 に示す．Meijin では，25 個が訳出できなかった．また，入力句 1 個に対し，出力句は，平均で 7〜8 件であった．

再現率と適合率を表 12 に示す．Meijin の再現率は，低いが，適合率は他の翻訳機より高かった．これは，正解 (○) が，複数解出た中に 1 つでもあればいいとしていることと，意味属性の制約が効いていることにあると考えられる．

また，Meijin の再現率が低かった理由は，今回作成したパターン辞書の作成に用いた標本が，[1] の重文・複文から抽出した名詞句のみであったためと考えられる．

表 11: 評価結果

	評価○	評価△	評価×
Meijin	74%(74 個)	1%(1 個)	25% (25 個)
ALT-J/E	87%(87 個)	12%(12 個)	1% (1 個)
翻訳の王様	94%(94 個)	5%(5 個)	1% (1 個)

表 12: 各翻訳機の性能

	再現率 R	適合率 P
Meijin	75%	99%
ALT-J/E	100%	87%
翻訳の王様	100%	94%

6　考察

実験で Meijin の訳出が無かった名詞句についての考察を 6.1 節に，評価が△，または，×となった名詞句について 6.2 節で考察する．

6.1　実験の考察

実験で Meijin の訳出が無かった名詞句の一部を以下に示す．

- このテーブルの位置
- 前科のある男
- 見え透いたうそ
- 世界中の少年たちの伝統的な夢
- 高い地位及び名声への道

この原因を以下に示す．

１．名詞句パターン辞書に日本語パターン自体が存在しない場合

（入力句１）高い地位および名声への道

この場合，名詞句パターン辞書の標本を増やすことにより解決できると考えられる．なお，入力句１に，類似する日本語パターンはなかった．

２．日本語パターン，および，英語パターンが存在するが，名詞意味属性制約で一致しない場合

（入力句２）湖の表面

（正解例）the surface of the lake

この入力句２に一番近いパターンは次の例である．

（日本語）N1(その他, 池) の N2(面, 表)

（英語）the N2 of the N1

この場合，入力句の “湖” とパターンの意味属性の “池”，および，入力句の “表面” とパターンの意味属性の “面, 表” は，単語の意味属性の距離が近い．そこで，名詞の汎化を考えることによって，パターンを適合できると考えられる．

6.2 パターン化の問題

以下にパターン化の誤り例を示す.
<元の句>

（日本語）あの力士

（英語）that sumo wrestler

<パターン>

（日本語）$REN1N2$(競技者, 男)

（英語）$PRN1$ sumo $N2$

本来，“力士＝ sumo wrestler” となる箇所が，今回単語アライメントを行ったことで “力士＝ wrestler” となっていた．パターン化対象の日本語表現において，“その力士” であり，英語表現が，“that wrestler” であれば，この対応でよかったが，今回の対象データでの英語名詞句では，“力士＝ sumo wrestler” となる必要がある．なぜならば，例えば，このパターンにマッチする例として，“あの俳優” があるが，この時に出力される英語名詞句は，“that smou actor” となり，明らかに違う訳出となるためである．

この問題を解決するためには，日本語名詞に対し，英語の単語をどこまで含むのかを検討しなければならない．最長一致などで，日本語名詞句と英語名詞句において，きちんと対応関係をとる必要がある．また，英語パターンにおいて，字面で残っている他のパターンに対しても同様である可能性があるため，見直す必要がある．

7　おわりに

近年，標本文には多くの線形要素が存在すること，文型パターン化作業の大半が機械化できることが分かり，大量のパターンを作ることが可能となったことで,「文型パターン」による翻訳が注目されている．最近では，重文・複文の日英機械翻訳にこの「文型パターン」を用いる手法が提案されており [1]．その手法では，パターンの記述要素に名詞句変数が使われ，その変数に代入された日本語表現を英訳する必要があった．

そこで，本研究では，大規模名詞句コーパスより，名詞句パターンを自動生成し，名詞句パターン辞書を作成した．そして， 名詞句翻訳プロトタイプシステム (Meijin) を用いて，名詞句パターン辞書の性能評価を行った．この結果は，再現率は 75%，適合率は約 99%であった．適合率とは，パターンの候補があった名詞句において，文法的，意味的に正しかった割合であり，本研究で作成した名詞句パターン辞書の有効性があったと考えられる．

今後の課題は，意味属性の汎化や名詞句パターンのアライメントの精密化，複合名詞への対応，および，標本を増やすことによる新たなルールの作成である．

謝辞

本論文作成に際して，多大なる検討と助言をして下さった池原教授ならびに村上仁一助教授，徳久雅人助手そして計算機 C 研究室の方々，に深く感謝します.

また，参考にさせて頂いた文献の著者の方々に対して感謝します.

参考文献

[1] 池原悟, 阿部さつき, 徳久雅人, 村上仁一: 非線型な表現構造に着目した重文と複文の日英文型パターン化, 自然言語処理, Vol.11, No.3, pp.69-95, 2004.

[2] 飯田朝子, 町田健: 数え方の辞典, 小学館, 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] Brill,E.: A simple rule-based part-of-speech tagger, ANLP-92, pp.152-155, 1992.

[5] 徳久雅人, 村上仁一, 池原悟: 文型パターンパーサの試作, 言語処理学会第 10 回年次大会発表論文集, pp.608-611, 2004.

付録１

名詞句パターン辞書の日本語パターンと
対応する英語パターン

表の見方についての説明

名詞句パターン辞書の日本語パターンについて頻度の高かった上位 74 位までを載せ，それに対応する英語パターンにおいて割合の高かった３位までとその他について載せる．

詳しい表の見方について説明する．まず，累計 45,675 件とは，名詞句コーパスから，パターン化した全件数である．次に,「日本語パターン」は上より，頻度の高かった第１位から順になっている．日本語パターンの第１位は，“REN1N2” の連体詞，名詞変数の連続であり，全件数の中で 3,719 件あった．

次に,「日本語パターンに対する英語の原文の種類数」を載せている．日本語パターン第１位の例では，英語の字面の種類数は，1,439 種類となる．

次に,「日本語パターンに対応する英語パターンの全種類数」を載せている．日本語パターンの第１位の例では，対応る英語パターンは 104 種類ある．

次に，英語パターンの頻度の高かった第１位，２位，３位とその件数を載せている．

最後の「残りのパターン数」とは，第１位，２位，３位の３種類を除いた英語のパターンの種類数である．日本語パターンの第１位では，“104 − 3” で残りのパターン数は 101 種類となっている．

最後に示した小さい表について説明する．累計が日本語パターンに対し，全件数の中で何件あったかを示す．累計 19 とは，ある日本語パターンに対し，19 件の原文があったこととなる．そして，原文で 19 件だった日本語パターンの種類数は１件である．累計 18 では，原文が 18 件だった日本語パターンの種類は４種類となる．

付録２

翻訳プロトタイプシステム (Meijin) による実験結果：
オープンテスト (100 件)

付録３

参考１（ALT-J/E）による実験結果：(100件)

付録４

参考２（翻訳の王様）による実験結果：(100 件)