概要

近年，翻訳方式として，文型パターンを用いた翻訳方法が注目されている．文型パターン翻訳では，翻訳対象となる原言語と目的言語の表現を文型パターンによって記述しておき，意味的に等価なパターンを対応づけることで，意味の失われない解析を実現しようとしている．この方式を実現するためには，あらかじめ，原言語と目的言語の対訳文から，意味的に非線形な構造を取り出して「対訳文型パターン辞書」を作成することが必要となる．ここで，言語表現の非線形要素とは，特定の概念を表現するための表現構造の要素のうち，他の要素に置き換えたら表現構造全体の意味が変わってしまうとき，その要素をその表現構造の「非線形要素」と言う．そして，１つ以上の非線形要素を有する表現構造を「非線形な表現構造」と言う．

文型パターン辞書構築の最大の問題は，文型パターンの網羅性と意味的排他性をいかにして実現するかの２点である．このうち，網羅性を実現するには，大量の文型パターンを汎化する必要がある．その汎化方法の１つとして，任意の文型要素が存在しても良い位置を表す「原文任意要素」と，文型パターン要素のうち，省略可能な要素を示す「文型任意要素」が文型パターンに追加されている．しかし，これらの任意要素指定機能は，実際，文型パターンの被覆率向上にどのくらいの効果をもたらしているか分かっていない．

そこで本研究では，任意要素指定機能による被覆率の向上の効果を評価することを目的とする．被覆率を評価するパラメータとしては,「文型再現率 $R1$」や,「平均適合パターン数 N」などが提案されているが，文型パターン辞書の規模がどのくらい拡大したかを直感的に測るために，本研究では，文型パターンの被覆率の向上の効果を推定するパラメータ η（文型パターン拡大率）を提案する．そして，η を用いて，任意要素指定機能の効果を評価する．　具体的な検証方法として，まず，現在使用されているパターン辞書（基準辞書）から，任意要素指定機能を除き，対象辞書を作成する．次に，文型パターンパーサを用いて，基準辞書および対象辞書のそれぞれとテスト用入力文を照合する．得られた照合結果から，各対象辞書の被覆率を，文型再現率 R1，および，平均適合パターン数 N を用いて求める．そして，被覆率の結果から，文型パターン拡大率 η を用いて，被覆率の向上の効果を推定する．

調査の結果，原文任意要素を文型パターンの記述に用いると，文型パターン辞書を 7〜20 倍に拡大したことに相当し，同様に，文型任意要素を文型パターンに用いると，文型パターン辞書を 1.4〜2.6 倍に拡大したことに相当することが分かった．よって，任意要素指定機能は，被覆率を大きく向上させる機能であることが定量的に分かった．

目 次

図 目 次

表 目 次

1　はじめに

言語表現には，言語の意味が表現構造と独立に扱うことができないという非線形の問題がある [1]．等価的類推思考の原理に基づく日英機械翻訳では，翻訳対象となる両言語の表現を「文型パターン」の対としておくことで，意味の失われない解析・生成を実現しようとしている [2]．文型パターンは，被覆率を向上させるため，様々な改良が施されてきた．その改良の１つとして，任意要素指定機能が文型パターンに追加されている．しかし，実際，被覆率の向上にどのくらいの効果がある機能なのか分かっていない．

そこで本研究では，任意要素指定機能による被覆率の向上の効果を評価することを目的とする．まず，文型パターンの被覆率の向上を推定するパラメータ η（文型パターン拡大率）を提案する．そして，現在使用されているパターン辞書（基準辞書）から任意要素指定機能を除き，η を実験的に求める．文型パターンに任意要素指定機能を加えたことで，文型パターン辞書の規模がどのくらい拡大したかを，η の低下具合から測り，効果を評価する．

本研究の構成は以下の通りである．2 章では，文型パターンと任意要素を説明する．3 章では，現在提案されている被覆率の評価パラメータと，本研究で提案する，文型パターンの被覆率向上を推定するパラメータ η について説明する．4 章では，任意要素指定機能の効果の調査方法と調査結果を示す．5 章では，実験の考察と，任意要素の付与基準の問題について述べる．

2　研究の背景

2.1　文型パターン辞書

2.1.1　文型パターンの概要

文型パターンとは，日英対訳標本文中の線形な要素を対象に，変数化，関数化，任意化，などを行ったもので，変数化された対象の文法的属性に着目して，単語レベル，句レベル．節レベルの 3 グループから構成される．ここで，線形要素とは，特定の概念を表現するための表現構造の要素のうち，他の要素に置き換えても表現構造全体の意味が変わらないとき，その要素を表現構造全体の「線形要素」という．

単語レベル文型パターンは，表現に使用される名詞，動詞などの自立語の線形な要素を変数化している．文法・単語レベルパターンの例を以下に示す．

文型パターンの例

- **日本語文：** ここから目黒へ行く間にとても静かな自然教育園があります。
- **日本語文型パターン:**/ytk ここから/tcfk$N1$へ/cf行く間に#2[/cf$ADV3$]/f$AJV4$!$N5$が/cf あります。
- **英文：**There is a very quiet nature study park between Meguro and here.
- **英語文型パターン：**There is #2[$ADV3$] $AJ4$ $N5$ between $N1$ and here.

上の例では，線形要素（目黒，Meguro）を”$N1$”,（とても，very）を” $ADV3$” というように変数化している．“ここから” や，“行く間に” は，は置き換えが不可能な要素なので，非線形要素として，変数化されずに字面でパターンに記述されている．文型パターンの N は名詞，ADV は副詞．AJV は形容動詞を表している．また，詳しくは次節で述べるが，/（小英字）は，原文任意要素の挿入を表し，#2[ ] は省略可能な要素を表している．

2.1.2 任意要素指定機能

文型パターンで使用される任意要素は,「原文任意要素」,「文型任意要素」に分けられる．以下にそれぞれの任意要素を示す.

(1) 原文任意要素

文型パターン記述において，原文任意要素は，文型パターン以外の要素の挿入位置を表すものであり，離散記号（/小英字）で指定される．該当する位置に現れた入力文の要素は，別途翻訳し，英文に組み込まなければならない[1]．原文任意要素として認める要素は，(1) 連用節，(2) 連体節，(3) 格要素，(4) 連用修飾要素，(5) 連体修飾要素の５種類とされている．原文文型任意要素として認める要素と挿入箇所を以下に示す.

- **連用節（/y）**
  節と節の間に連用節の挿入を認める．同一の節内には挿入を認めない．離散記号 /y で示される.

- **連体節（/t）**
  名詞句の直前に連体節の挿入を認める ．同一の名詞句内には挿入を認めない．離散記号 /t で示される.

- **格要素（/c）**
  文型パターン内に存在する格要素の前後に別の格要素の挿入を認める．但し，同一格要素内に別の格要素が挿入されてはならない．また，格要素のないところに格要素を挿入してはならない．離散記号 /c で示される.

- **連用修飾要素（/f）**
  挿入可能な連用修飾要素は，形容詞連用形，副詞，副詞句のいずれかとする．挿入可能な位置は，文頭，述部の直前，形容詞の直前，動詞の直前とする．離散記号 /f で示される.

- **連体修飾要素（/k）**
  挿入可能な連体修飾要素は，連体詞,「A の」・「A と」などの名詞句構成要素，形容詞連体形，動詞連体形のいずれかとし，文型パターン内の名詞句直前への挿入を認める．同一名詞句内への挿入は認めない． 離散記号 /k で示される.

[1]日本語文型パターンに対応する英語文型パターンには，原文任意要素の挿入位置が記されていないので，英文に組み込む際に，構文解析処理などで組み込む位置を知る必要がある.

以下に離散記号を挿入した文型パターンの例を示す.

- **日本語文**
  その知らせを聞いて彼女の顔は明るくなった。

- **文型パターン**
  /y1GEN$1/k$N$2 を /cf$V$3(て|で) $1^{#4[/tk$N$5 の]}/k$N$6 は /cf 明るくなった。
  1 2 3 4 2 3

  1. 文型の先頭に連用節 (/y) の挿入を認める
  2. 名詞の直前に連体修飾要素 (/k) の挿入を認める
  3. 格要素の直後に格要素 (/c) の挿入を認め，また，動詞の直前に連用修飾要素 (/f) の挿入を認める
  4. 名詞の直前に連体修飾要素 (/k) の挿入を認め，また，名詞句の直前に連体節 (/t) の挿入を認める

(2) 文型任意要素

文型任意要素は，文型パターン要素のうち，省略可能な要素を示す．それが削除されても英語文型自体は変化しないが，それ自身の訳語選択の決定や訳語挿入位置の決定が困難であるなど，要素自身の翻訳に困難さが生じる要素が，文型任意要素として，任意要素記号（[]）で指定されている．文型任意要素は，英文中に訳出すべき位置情報が指定されているため，該当する部分の翻訳を組み込むことは容易である[2].

以下に任意要素記号を用いた文型パターンの例を示す.

- **日本語文**
  彼女の歌を聞きながら聴衆は手拍子で合いの手を入れた。

- **文型パターン**
  /ytk#1[N2 の]/kN3 を/cf$V$4 ながら/tk$N$5 は/tcfk 手拍子で/tcfk 合いの/k 手を/cf 入れた。
  日本語文の省略可能な要素（" 彼女の"）が，文型任意要素として文型パターンに指定されている.

[2]例えば，2.1.1 節の文型パターンの例では，日本語文型パターンにおいて，文型任意要素として示されている#2[/cfADV3] が，英語文型パターンにも記されている.

2.1.3 文型パターン辞書

文型パターン辞書は，日本語の基本的な表現が収録されていると見られる辞書や，語学教育用の教科書，機械翻訳機能評価用の試験文等，約 30 種類のドキュメント（日英対訳文 100 万件）から，２箇所，又は３箇所の述部を持つ日本語の重文（文接続のある文），複文（埋め込み文のある文）の対訳標本文約 12.9 万件を取り出し，それを汎化することによって作成されており，汎化の程度により，単語レベル，句レベル，節レベルの３種類の文型パターン（異なり文型パターン 22.1 万件）が収録されている．

単語レベルの文型パターン辞書（異なり文型パターン 12.3 万件）では，原文任意要素が約 70 万回挿入されており，１文型パターンあたり，平均６箇所に離散記号が挿入されている．また，文型任意要素は約４万回挿入されており，１文型パターンあたり，平均 0.3 回文型任意要素記号が挿入されている．

2.1.4 文型パターンの照合

入力文に適合する文型パターン，および，適合の仕方をすべて出力するプログラムとして，文型パターンパーサがある [3].

入力文の翻訳に使用できる文型パターンは，必ずしも入力文のすべての要素が適合する文型パターンである必要はなく，入力文の主要な構造が適合し，意味的に正しいパターンであればよい．適合文型は，以下の 2 種類に分類できる．

- **完全一致文型：**入力文のすべての要素が文型パターンの要素と適合する文型パターン
- **部分一致文型：**入力文の一部の要素が原文任意要素であり，離散記号と適合する文型パターン

以下に入力文に適合した「完全一致文型」と「部分一致文型」の例を示す.

- **入力文：**彼は頭がいい上に、勉強家である。
- **完全一致するパターン：**/y/tk$N1$ は/cfk$N2$ が/cf$AJ3$ˆrentai!上に、/tk$N4$.da。
  - $N1$ = 彼
  - $N2$ = 頭
  - $AJ3$ = いい
  - $N4$ = 勉強家

- **部分一致するパターン：** /y#1[/tk 非常に/cf$AJV2$^rentai]!$N3$ は/tcfk$N4$.da。
  （原文任意要素の対応先を<< >>の記号で表す.）
  - $N3$ = 彼
  - /tcfk = <<頭がいい上に>>
  - $N4$ = 勉強家

2.2 被覆率の評価パラメータ

[4] では，被覆率を評価するパラメータとして,「文型再現率 $R1$」,「平均適合パターン数 N」などを提案している.

2.2.1 文型再現率 $R1$ と平均適合パターン数 N

文型再現率 $R1$ は，入力文に対して適合文型パターンが存在するかどうかを文単位で集計したもので，下式で定義される.

$$\text{文型再現率 } R1 = M/I$$

（但し，I:テスト用入力文の数 M:「自己パターン」以外の適合文型パターンが 1 つ以上存在した入力文の数）

平均適合パターン数 N は，入力文に対して，完全一致あるいは部分一致する文型パターン数の平均値を表し，下式で定義される.

「完全一致文型数」の平均

$N1$ = 完全一致文型パターン数 / 入力文の数

「部分一致文型数」の平均

$N2$ = 部分一致文型パターン数 / 入力文の数

「一致文型数」の平均

$N = N1 + N2$

2.2.2 被覆率の求め方

下図1を用いて，文型再現率 $R1$，および，平均適合パターン数 N の求め方を示す．まず，入力文１～３と文型パターンを照合し，適合するパターンを抽出する．入力文１には適合パターンが２つ存在し，入力文２は適合パターンが存在せず，入力文３には適合パターンが１つ存在した．入力文３文中，２文は適合パターンが存在するので，この場合，R 1 は，$R1 = 2/3 = 66.7(\%)$ となる．また，適合パターンが合計３つ存在するので，N は $N = 3/3 = 1$(パターン) となる．

文型再現率 $R1$，および，平均適合パターン数 N は，文型パターン辞書の規模に依存しているため，文型パターン辞書の規模が大きいほど，$R1$ および N の値が高くなる．

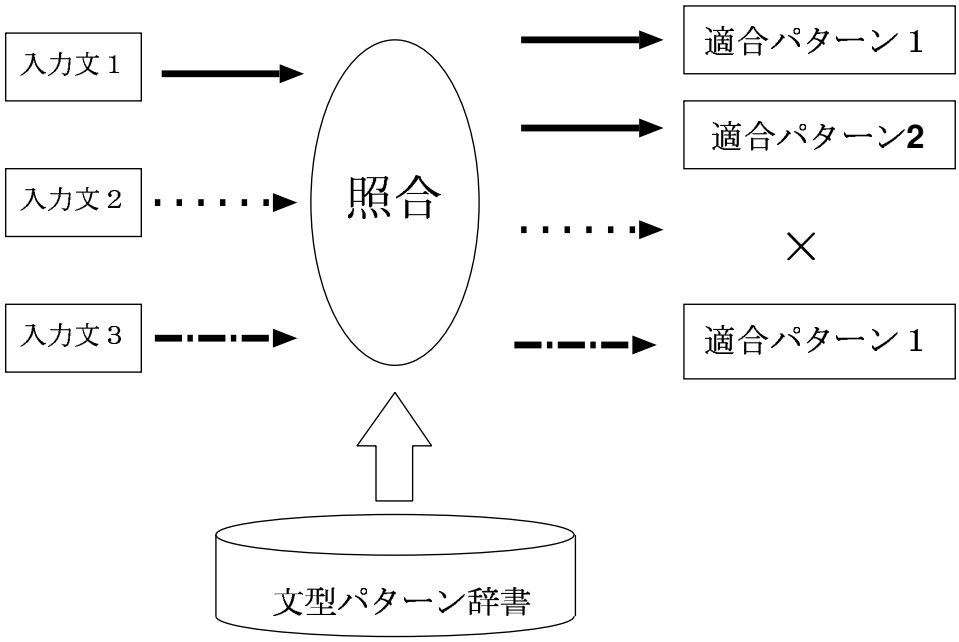

図 1: $R1$ と N の算出方法

2.2.3 被覆率の評価パラメータと文型パターン数の関係

文型パターン数と，文型再現率 $R1$ の関係を図2に，また．文型パターン数と平均適合パターン数 N の関係を図3に示す．

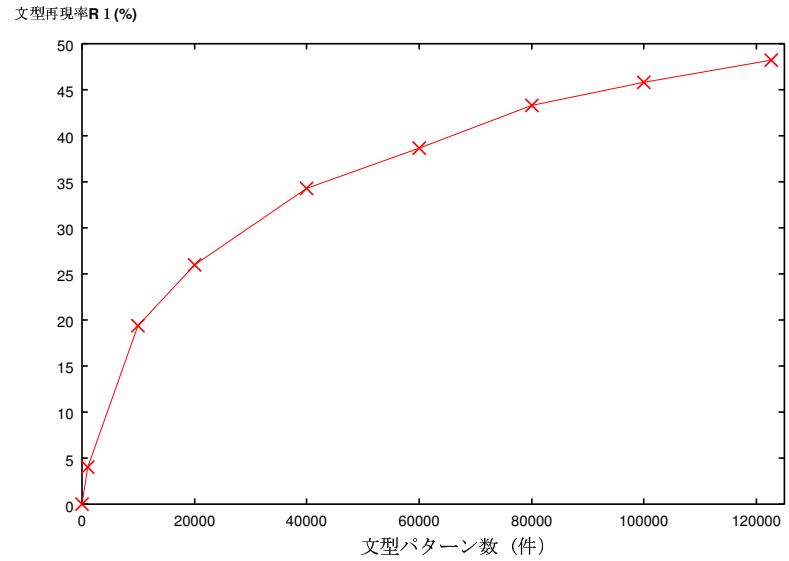

図 2: $R1$ と文型パターン数の関係

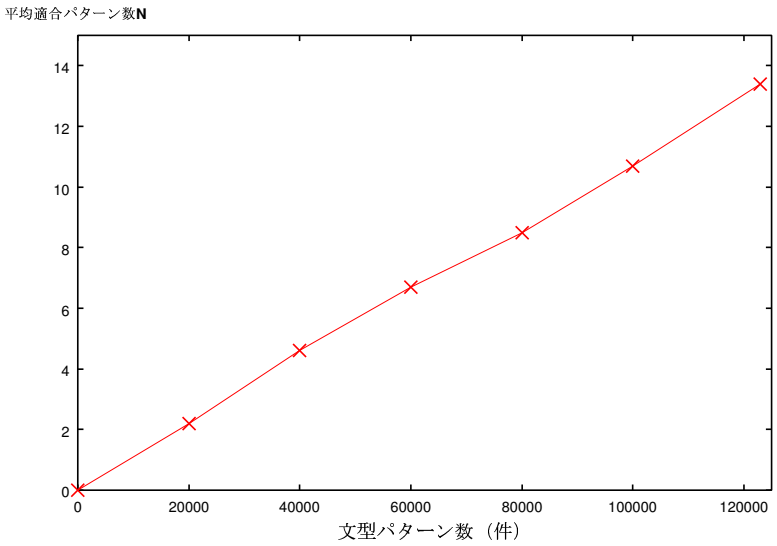

図 3: N と文型パターン数の関係

図より，文型再現率 $R1$ は，文型パターン数が多くなると，飽和傾向が見られた．また，平均適合パターン数 N は，文型パターン数と比例関係にあることが分かった．

3　被覆率向上の効果の推定パラメータ

3.1　文型パターン拡大率 η

本研究では，被覆率の向上を推定するパラメータとして，文型パターン拡大率 η を提案する．η は,「評価対象の文型パターン辞書（対象辞書）が，基準文型パターン辞書（基準辞書）の文型パターン数に換算して，何倍に相当するか」を表し，以下の式で定義する．

$$\text{文型パターン拡大率}\eta = X/B \tag{1}$$

（但し，B:基準辞書の文型パターン数 ，X:対象辞書の文型パターン数の換算値）

3.2　換算による文型パターン拡大率 η の算出

ここで，対象辞書の文型パターン数の換算値 X の換算方法が問題となる．被覆率の評価パラメータに基づき換算値を求めることにすると,「対象辞書の被覆率に至るには，基準辞書のパターン数をどれだけにすれば良いか」という換算ができる．したがって，基準辞書の被覆率を文型パターン数 p の関数 $c(p)$ で表しておき，対象辞書の被覆率を r として，$r = c(X')$ を満たす X' を η を求める際に用いる．さらに，一般に関数 $c(p)$ は近似曲線であり，誤差を含む．そこで，基準辞書の η を 1 にするために，η の計算において，実際の値である B を用いるのではなく，基準辞書の被覆率を R として，$R = c(B')$ を満たす B' を用いることにする．

したがって，近似曲線 c で換算した文型パターン拡大率 η_c は以下の式で求める．

$$\eta_c = c^{-1}(r)/c^{-1}(R) \tag{2}$$

（但し，R:基準辞書の被覆率，r:対象辞書の被覆率，c^{-1}:近似曲線の逆関数であり，被覆率に対する文型パターン数）

3.2.1　文型再現率 $R1$ による文型パターン拡大率 η_{R1} の求め方

基準辞書（本研究では，文法・単語レベルパターン辞書（122,619 パターン収録）を使用する）を用いた実験から，文型再現率 $R1$ と文型パターン数 p の関係を，図 4 で示す．

図中の縦軸は，文型再現率 $R1$ を示し，横軸は文型パターン数 p を示す．サンプル点（×）は，文型パターン数と $R1$ の実測値を示す．

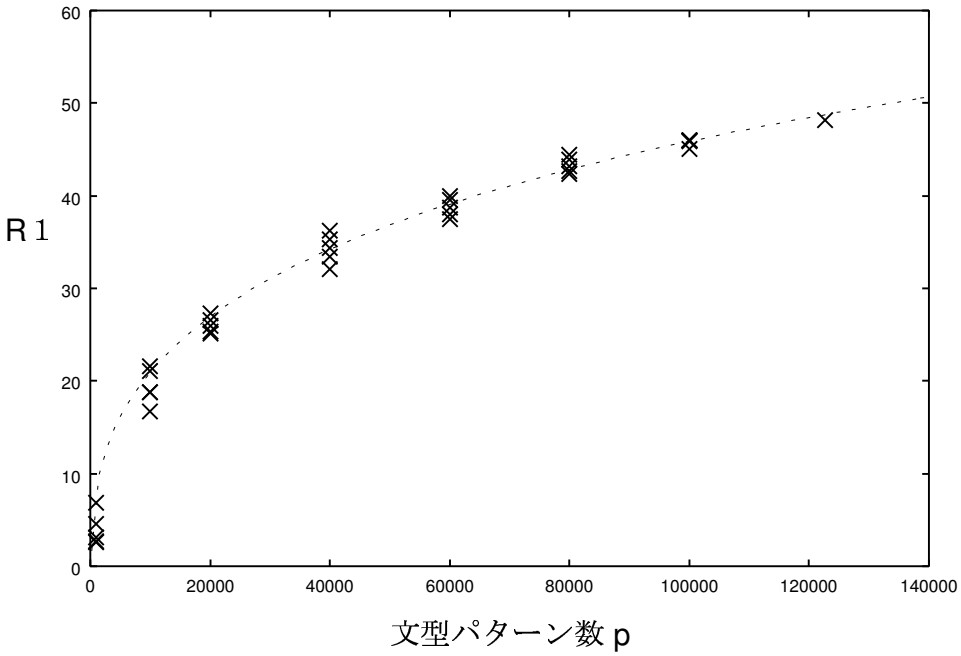

図 4: R1 と文型パターン数の関係図

非線型回帰分析より，文型再現率 $R1$ の文型パターン数 p に対する特性を，(3) 式で近似する．近似曲線を図中に点線で示す．

$$R1 = (1 - exp(-\lambda_1 p^{\lambda_2})) \times 100 \qquad (\%) \tag{3}$$

（但し，非線形回帰分析より，$\lambda_1 = 0.005038$，$\lambda_2 = 0.4171$）

実験で得られた被覆率を近似式に代入することによって得られる p を，対象辞書の文型パターン数の換算値 X とする．そして，(1) 式より，文型再現率 $R1$ による文型パターン拡大率 η_{R1} が求まる．

3.2.2 平均適合パターン数 N による文型パターン拡大率 η_N の求め方

基準辞書を用いた実験から，平均適合パターン数 N と文型パターン数 p の関係を，図 5 に示す．図中の縦軸は，平均適合パターン数 N を示し，横軸は文型パターン数 p を示

す．サンプル点（×）は，文型パターン数と N の実測値である．

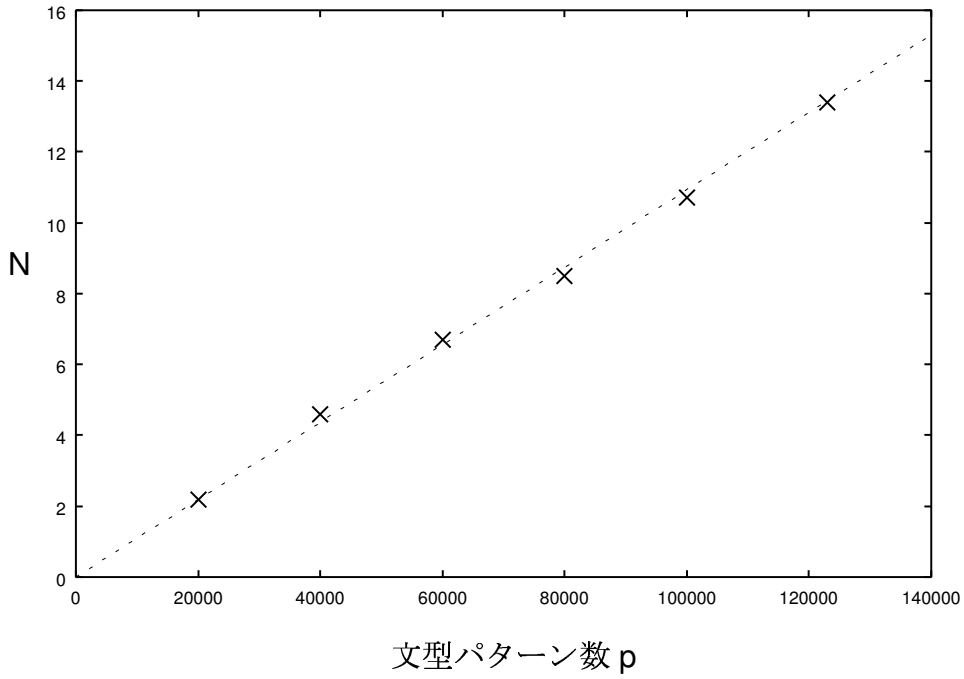

図 5: N と文型パターン数の関係図

線形回帰分析より，文型再現率 N の文型パターン数 p に対する特性を，(4) 式で近似する．近似線を図中に点線で示す．

$$N = \lambda_3 p \tag{4}$$

（但し,$\lambda_3 = 13.4085/122,619$）

同様にして，(1) 式より，平均適合パターン数 N による文型パターン拡大率 η_N が求まる．

3.2.3 パターン記述による文型パターン拡大率 η_d の求め方

任意要素などの記号や，関数を使用したパターンは，何通りかの解釈ができる．その解釈の数を用いて，文型パターン拡大率を求める．

<例>

- P1：/y$N1$ は $N2$ を V3 て#4[$N5$ の]$N6$ を $V7$
  - P1-1：$N1$ は $N2$ を $V3$ て $N6$ を $V7$
  - P1-2：/y$N1$ は $N2$ を V3 て $N6$ を $V7$
  - P1-3：$N1$ は $N2$ を V3 て $N5$ の $N6$ を $V7$
  - P1-4：/y$N1$ は $N2$ を V3 て $N5$ の $N6$ を $V7$

上の例では，文型パターン P1 は P1-1〜4 の，４通りの解釈が出来る．このように，対象辞書の解釈の数を算出し，その値を対象辞書の文型パターン数の換算値 X として，(1) 式よりパターン記述による文型パターン拡大率 η_d を求める．

この方法による η の算出は [5] により，検討している．

4　任意要素の効果の調査

4.1　調査対象と目的

本調査では，2.2 節で述べた,「原文任意要素 5 種類，および，文型任意要素」の有無と被覆率の関係について調査する．

そこで，4.2 節の方法で，基準辞書から，それぞれの任意要素指定機能を削除し，対象辞書を作成する．(i) 原文任意要素を全て削除した辞書，(ii) 原文任意要素・連用節 (/y) を削除した辞書，(iii) 原文任意要素・連体節 (/t) を削除した辞書，(iv) 原文任意要素・格要素 (/c) を削除した辞書，(v) 原文任意要素・連用修飾要素 (/f) を削除した辞書，((ii) 原文任意要素・連用節 (/y) を削除した辞書，(vii) 文型任意要素を削除した辞書の計 7 種類の対象辞書を作成し，これらを調査対象とする．

4.2　対象辞書作成方法

原文任意要素，および，文型任意要素のそれぞれの効果を調査するために，基準辞書から各要素を除いた対象辞書を作成する．

例：基準パターンから，(1)「原文任意要素・連体節 (/t)」と (2)「文型任意要素」を削除した文型パターン

- **日本語文：**こんなに客が少なくては商売上がったりだ。
- **基準パターン：**/y#1[こんなに]/tk 客が/cf$AJ2$(て|で) は/tcfk$N3$ 上がったりだ。
- **(1) を削除：**/y#1[こんなに]/k 客が/cf$AJ2$(て|で) は/cfk $N3$ 上がったりだ。
- **(2) を削除：**/y こんなに/tk 客が/cf$AJ2$(て|で) は/tcfk$N3$ 上がったりだ。

このような方法で，任意要素別に対象辞書を作成する．

4.3　調査方法

文型パターンパーサを jpp 用いて，テスト用入力文と 7 種類の対象辞書，および，基準辞書と照合する．その結果から，各対象辞書の被覆率を評価し，そして，被覆率向上を推定する．

テスト用入力文としては，文型パターンの作成に使用された対訳標本の日本文（123,016 文）を使用する．

4.4 調査の様子

具体的な実験の手順を以下に示す.

1. 文型照合実験に必要なデータベース，および，プログラムを準備する

   - データベース：入力文の形態素解析結果（ALL.morph）

     入力文：彼のお母さんがああ若いとは思わなかった。
     形態素解析結果
     INPUT=AC000004-00=0x7ffff=0xfff
     彼のお母さんがああ若いとは思わなかった
     1．/彼 (1710,{NI:23,NI:48,KR:0601s10})
     2．+の (7410)
     3．/お母さん (1100,{NI:80,NI:49})
     4．+が (7410)
     5．/ああ (1110,{NI:1235,KR:2907z07})
     6．/若い (3106,{NY:5,KR:8803c00})
     7．+と (7420)
     8．+は (7530)
     9．/思わ (2392, 思う,{NY:32,NY:31,KR:0601a01})
     10．+なかっ(7184, ない)
     11．+た (7216)
     12．+。([P]0110)
     13．/nil

   - 文型照合プログラム（パターンパーサ）

2. 対象辞書および基準辞書のオートマトン（マップファイル）を作成する.

   % pattern bwjall（辞書ファイル名）
   → map.bwjall（マップファイル）

3. 照合

   % matching map.bwjall s1(形態素ファイル名)
   → s1.result（結果ファイル）

4. 出力の例
   入力文：この俳句は冬の景色を読んだものだ。

   形態素ファイル：AF029183-00

   候補パターン

   WJ127171-01:/y#1[GEN2]/kN3 は/tcfkN4 を/cfV5.kako^rentai!ものだ。
   WJ188961-01:/y#1{#1[GEN2]/kN3 は,/tcfkN4 を}/cfV5.kako^rentai!ものだ。

結果ファイル例

INPUT=AF029183-00=
この俳句は冬の景色を読んだものだ。
1. /この (4200)
2. /俳句 (1100,{NI:1042,KR:1908k24})
3. +は (7530)
4. /冬 (1500,{NI:2676,KR:8508x00})
5. +の (7410)
6. /景色 (1100,{NI:525,KR:0403i03})
7. +を (7430)
8. /読ん (2374, 読む,{NY:23,NY:32,NY:26,KR:1508a36})
9. +だ (7227)
10. /もの (1800,{NI:1})
11. +だ (7256)
12. +。([P]0110)
13. /nil
—
PATTERN=WJ127171-01=/y#1[GEN2]/kN3は/tcfkN4を/cfV5.kako^rentai!ものだ。
=[/k,N3, は,/tcfk,N4, を,V5,.kako,!, ものだ。]=0x1ff9c=
N3=俳句=俳句=n(napf(2))=0x0000c=
N4=景色=景色=n(napf(6))=0x00180=
V5=読む=読ん=v(8)=0x00c00=
—
PATTERN=WJ188961-01=/y#1{#1[GEN2]/kN3は,/tcfkN4を}/cfV5.kako^rentai!ものだ。
=[/k,N3, は,/tcfk,N4, を,V5,.kako,!, ものだ。]=0x1ff9c=
N3=俳句=俳句=n(napf(2))=0x0000c=
N4=景色=景色=n(napf(6))=0x00180=
V5=読む=読ん=v(8)=0x00c00=
#1=[N3, は,N4, を]=0x0039c=
#1=[N3, は,N4, を]=0x0039c=
PATTERN=WJ188961-01=/y#1{#1[GEN2]/kN3は,/tcfkN4を}/cfV5.kako^rentai!ものだ。
=[GEN2,N3, は,/tcfk,N4, を,V5,.kako,!, ものだ。]=0x1ff9f=
GEN2=この=この=gen(1)=0x00003=
N3=俳句=俳句=n(napf(2))=0x0000c=
N4=景色=景色=n(napf(6))=0x00180=
V5=読む=読ん=v(8)=0x00c00=
#1=[GEN2,N3, は,N4, を]=0x0039f=
#1=[GEN2]=0x00003=

—

結果ファイルの構成は，以下のようになっている．

- INPUT=「入力形態素ファイル名」
- 入力文
- 形態素情報
- PATTERN=「適合パターン番号」=「パターン記述」=「経路情報」=「適合文字位置ビット」
- 変数名=「標準形表記」=「原文表記」=「経路情報」=「適合文字位置ビット」

5. 文型再現率 $R1$ の算出
照合結果ファイルより，適合文型パターンが抽出された入力文の数を算出する．そして，2.2.1 節の方法によって $R1$ を求める．なお，入力文から作成されたパターン「自己パターン」は，適合文型パターンには含めないこととする．

6. 平均適合パターン数 N の算出
照合結果ファイルより．抽出された適合文型パターンの数を算出する．ここで，適合文型パターン数の計算において，注意点が 2 点ある．1 つ目は，$R1$ の算出の際と同様に，自己パターンは含めないことである．2 つ目は，適合解の数え方である．適合解には変数のバインド情報が含まれている．同一パターンで複数の適合解がある場合，各変数や関数の対応する形態素が適合解ごとに異なることで，適合解の個数を数える．(例えば，4. 結果ファイル例の場合，適合文型パターン数は，3 パターンとなる．)

7. 文型パターン拡大率 η の算出
求めた $R1$ と N の値を用いて，3.2.1 節，および，3.2.2 節の方法で η を算出する．

4.5 調査結果

照合結果より，各評価パラメータの値を表 1 にまとめる．

表 1: 調査結果

使用するパターン辞書	$R1$ による推定		N による推定	
	$R1(\%)$	η_{R1}	N	η_N
基準辞書	48.2	1.0	13.4	1.0
(i) 原文任意要素すべて削除	17.0	0.05	1.9	0.14
(ii) 連用節（/y）削除	42.1	0.65	10.9	0.81
(iii) 連体節（/t）削除	47.6	0.96	13.1	0.98
(iv) 格要素（/c）削除	39.2	0.51	6.8	0.50
(vi) 連用修飾要素（/f）削除	44.6	0.77	11.9	0.89
(v) 連体修飾要素（/k）削除	45.6	0.83	11.1	0.83
(vii) 文型任意要素すべて削除	35.6	0.39	9.7	0.72

(i) の文型再現率 $R1$ による文型パターン拡大率 η_{R1} および平均適合パターン数 N による文型パターン拡大率 η_N より，原文任意要素を文型パターンの記述に用いると，$R1$ から推定すると，基準辞書を約 20 倍（$1.0/0.05 = 20$）拡大したことに相当し，N から推定すると，約 7 倍（$1.0/0.14 \approx 7$）に拡大したことに相当する．同様に，(vii) の η_{R1}，η_N より，文型任意要素を文型パターンの記述に用いると，$R1$ から推定すると，基準辞書を約 2.6 倍拡大したことに相当し，N から推定すると，約 1.4 倍に拡大したことに相当する．

5　考察

5.1　任意要素の効果

調査結果より，任意要素指定機能は被覆率を大きく向上させることが分かった．原文任意要素をすべて使用する時，または，文型任意要素をすべて使用する時，文型再現率 $R1$ による文型パターン拡大率 η_{R1} と平均適合パターン数 N による文型パターン拡大率 η_N に差がある．これは，N は，入力文当たりの適合パターン数を測り，$R1$ は適合パターンのある入力文を測るため，適合パターンのある入力文の数が，増大しており，特定の入力文において，適合パターン数が多くなっているのではないと思われる．

つまり，$\eta_{R1} > \eta_N$ の場合，適合パターンのある入力文を増やす効果が高く，$\eta_N > \eta_{R1}$ の場合，特定の入力文において，適合パターンを増やす効果が高いと言えると考えられる．

5.2　原文任意要素の挿入要素別の効果

原文任意要素の挿入要素別の結果を見ると，連体節（/t）はあまり効果がなさそうである．連体節（/t）で，あまり効果が出なかった理由を考察する．

離散記号は，2.2.1 節のように挿入箇所が決められており，同じ位置に複数の離散記号が挿入されている．よって，文型パターンに，/tcfk となっているところがあるが，そこで連用節（/t）が使用されていなくても，格要素（/c），連体修飾要素（/k），連用修飾要素（/f）があれば，連体節が挿入可能になることがある．そのため，連体節（/t）の効果があまり出なかったと思われる．以下に例を示す．

- **入力文：**この本は遠藤氏が新聞に連載したコラムをまとめたものだ。
- **適合したパターン：**/y#1[$GEN2$]/k$N3$ は/tcfk$N4$ を/cf$V5$.kako^rentai!ものだ。
  - /k = この
  - $N3$ = 本
  - /tcfk = <u>遠藤氏が新聞に連載した</u> /t
  - $N4$ = コラム
  - $V5$ = まとめ
- **/t を除いたパターン：**/y#1[$GEN2$]/k$N3$ は/cfk$N4$ を/cf$V5$.kako^rentai!ものだ。

- /k = この
- $N3$ = 本
- /cfk =遠藤氏が 新聞に 連載した
  /c /c /k
- $N4$ = コラム
- $V5$ = まとめ

5.3 意味的排他性の評価

任意要素を用いた文型パターンが，意味的に正しい文型パターンであるか，また，それに対応する英語パターンが訳文の生成に問題なく使用できるかを調査した.

5.3.1 原文任意要素

原文任意要素の挿入要素により，意味的に不適切なパターンと適合する割合にあまり違いはなかった．しかし，離散記号と適合した入力文の要素が，英語翻訳の際の重要な語になっている場合，もしくは，入力文のほとんどの要素が原文任意要素となる場合，得られたパターンは単純なものであったり，意味的に正しくないパターンである場合が多かった.（適合パターンのうち，約４割が意味的に正しくないパターンだと思われる.） 以下に例を示す.

例１

- **日本語文：**皆いっせいに手を上げたので、先生は誰を当てたらよいか迷った。
- **英文：**They raised their hands all at once, so the teacher did not know whom he should call on.
- **適合した日本語パターン**
  /y$1^{#1[$GEN2$]}/k$N3$ は/cf$V4$.kako^katei$1/f$V5$.kako。
  - /y = 皆いっせいに手を上げたので、
  - $N3$ = 先生は
  - /cf = 誰を

- $V4$ = 当てる
- /f = 良いか
- $V5$ = 迷う

- **適合した日本語パターンに対応する英語パターン：**#1[$AJ2$] $N3$ $V5$.past in V4

例１の例文の，英語翻訳の際の重要な語となる,「ので」という部分と「よいか」という部分が原文任意要素と対応している．そのため，得られた日本語パターンに対応する英語パターンは，意味的に正しくないパターンになっている．

例２

- **日本語文：**話に気を取られて我れ知らず乗り出して聞いていた。
- **英文：**The story interested me so much that I was unconsciously leaning forward to listen.
- **適合した日本語パターン**
  /y</tk$N1$ は>#2[!熱狂的な (ほど|程)]/cf ($V3.teiru.kako$|$ND3$ をしていた)。
  - /y = 話に気を取られて我れ知らず乗り出して
  - V3 = 聞く
- **適合した日本語パターンに対応する英語パターン:**$N1$ be.past #2[frantically]($V3$|$V(ND3)$).

例２では，日本語文の「話しに気を取られて我知らず乗り出して」の部分が原文任意要素と対応している．日本語文の要素のほとんどが，原文任意要素となっているので，得られた日本語パターンに対応する英語パターンは，単純である．日本語文の要素のうち，原文任意要素と対応する要素が多い程，得られたパターンは意味的に正しくないパターンである可能性が高いと思われる．

このような問題を解決するためには，離散記号の付与基準や，文型パターンの照合条件を見直す必要がある．

5.3.2 文型任意要素

任意要素記号を挿入したパターンは，実際の翻訳の際に使用できるかどうかを調査した．その結果，適合したパターンの中には，意味的に正しくないパターンも含まれてい

た．しかし，任意要素を使用したことが原因で，意味的に正しくないパターンになっているものはなかった．なお，任意要素記号は，その部分が削除されても英語文型自体は変化しないものに付与されているので，任意要素が使われているパターンと適合しても問題はない思われる．

6　おわりに

本研究では，まず，文型パターンの被覆率向上の効果を推定するパラメータとして，文型パターン拡大率 η を提案した．任意要素を用いた効果を，文型再現率 $R1$，および，平均適合パターン数 N から推定した η を用いて評価した．具体的には，現在使用されている文型パターン辞書のパターン数と，$R1$，および，N の関係式を実験的に求め，得られた関係式を用いて，η を算出した．調査結果より，原文任意要素を文型パターンの記述に用いると，$R1$ から η を推定すると，文型パターン辞書の規模を 20 倍に拡大したことに相当し，また，N から η を推定すると，約 7 倍に拡大したことに相当することが分かった．同様に，文型任意要素を文型パターンの記述に用いると，$R1$ から η を推定すると，文型パターン辞書の規模を約 2.6 倍に拡大したことに相当し，N から η を推定すると，約 1.4 倍に拡大したことに相当することが分かった．よって，任意要素の指定は，被覆率の向上に大きな効果をもたらしていると言える．

しかし，原文任意要素を使用したパターンの意味的排他性が低いので，任意要素の記述方法や挿入箇所などを見直す必要がある．

謝辞

本論文作成に際して，多大な検討と助言をしてくださった池原悟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工学 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, 阿倍さつき, 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, 11(3), pp.69-95, 2004.

[2] 池原悟, 佐良木昌, 宮崎正弘, 池田尚志ほか:等価的類推思考の原理による機械翻訳方式, 電子情報通信学会技術研究報告, TL2002-34, pp.7-12, 2002.

[3] 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.

[4] 池原悟, 徳久雅人, 村本奈央:日本語重文・複文を対象とした文法レベル文型パターンの被覆率特性, 自然言語処理, 11(4), pp.147-178, 2004.

[5] 金澤佑哉, 徳久雅人, 村上仁一, 池原悟:文型パターンにおける時制・相・様相表現の汎化とその効果, 言語処理学会第 11 回年次大会, 2005(発表予定).

[6] 遠藤久美子, 徳久雅人, 村上仁一, 池原悟:文型パターンにおける任意要素の記述方法とその効果, 言語処理学会第 11 回年次大会, 2005(発表予定).