

# 012006 文型パターンにおける任意要素指定の効果

計算機工学講座 池原研究室 遠藤久美子

## 1 はじめに

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。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。任意要素指定機能は、その1つである。

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。そこでまず、文型パターンの被覆率の向上を推定するパラメータ  $\eta$  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くことで、 $\eta$  の低下の具合を測り、その機能の効果を評価する。

## 2 評価パラメータ

### 2.1 被覆率の評価パラメータ

[?]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。 $R1$ は、多くの文を入力した際、文型パターンの適合する文がどれだけあるのかを表す。 $N$ は、入力文に対して適合する文型パターンがどれだけあるのかを表す。それぞれ、以下の式で定義される。

$$\text{文型再現率 } R1 = M/I \quad (1)$$

$$\text{平均適合パターン数 } N = P/I \quad (2)$$

(但し、 $I$ :テスト用入力文の数、 $M$ :適合文型パターンが1つ以上存在した入力文の数、 $P$ :適合文型パターン数)

### 2.2 被覆率向上の効果の推定パラメータ

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  $\eta$  を提案する。 $\eta$ 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。

$$\text{文型パターン拡大率} = X/B \quad (3)$$

(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)

$X$ は次の2つの方法で求める。

#### (1) $R1$ による文型パターン拡大率 $\eta_{R1}$ の求め方

基準辞書を用いた実験から、 $R1$ の文型パターン数  $p$  に対する特性を、次式で近似する。

$$R1 = 100(1 - \exp(-\lambda_1 p^{\lambda_2})) \quad (4)$$

(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、 $X$  とする。そして、(3)式より、 $\eta_{R1}$  が求まる。

#### (2) $N$ による文型パターン拡大率 $\eta_N$ の求め方

同じく、 $N$ の特性を、次式で近似する。

$$N = \lambda_3 p \quad (5)$$

(但し、 $\lambda_3 = 13.4085$ )

同様に、(3)式より、 $\eta_N$  が求まる。

## 3 任意要素の効果の調査

### 3.1 調査対象と目的

文型パターン記述において、原文任意要素は、ある条件を満たす表現の挿入可能位置を表す。具体的には、(1)連用節(/y)、(2)連体節(/t)、(3)格要素(/c)、(4)連用修飾要素(/f)、(5)連体修飾要素(/k)である。

文型任意要素は、文型パターン要素のうち、省略可

能な要素を表す。現在使用されている、約12万パターンが収録された文法・単語レベルパターン辞書(本研究の基準辞書とする)[?]では文型任意要素が6,871種類あり、合計で39,531回使用されている。

本調査では「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。

### 3.2 パターン辞書作成方法

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。

基準パターンの例

/ytkTIME1に#2[/cfADV3]/cfV4とは/cfAJV5。

/fを除いたパターンの例

/ytkTIME1に#2[/cADV3]/cV4とは/cAJV5。

### 3.3 調査方法

3.2節で作成した7種類の対象辞書、および、基準辞書をそれぞれ用いて、テスト用入力文(123,016文)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。

### 3.4 調査結果

各評価パラメータの値を表1にまとめる。

(i)の $\eta_{R1}$ および $\eta_N$ より、原文任意要素を文型パターンの記述に用いると、基準辞書を7~20倍に拡大したことに相当し、同様に、(vii)の $\eta$ より、文型任意要素の効果は1.4~2.6倍相当であることが分かった。

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	
	$R1(\%)$	$\eta_{R1}$	$N$	$\eta_N$
基準辞書	48.2	1.0	13.4	1.0
(i) 原文任意要素すべてなし	17.0	0.05	1.9	0.14
(ii) /y 除く	42.1	0.65	10.9	0.81
(iii) /t 除く	47.6	0.96	13.1	0.98
(iv) /c 除く	39.2	0.51	6.8	0.50
(vi) /f 除く	44.6	0.77	11.9	0.89
(v) /k 除く	45.6	0.83	11.1	0.83
(vii) 文型任意要素すべてなし	35.6	0.39	9.7	0.72

## 4 考察

対象辞書により、 $\eta_{R1}$ と $\eta_N$ に差がある。それは、 $N$ と $R1$ の測る値の性質より、特定の入力文において適合パターン数が高くなるのではなく、適合パターンのある入力文の数が、増大しているためと思われる。

## 5 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  $\eta$  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  $\eta$  を用いて評価した。調査結果より、任意要素の指定は、被覆率を大きく向上させることが分かった。

## 参考文献

- [1] 徳久ほか: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.
- [2] 池原ほか: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.
- [3] 池原ほか:非線形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, 言語処理学会論文誌, 11(3), pp.69-95, 2004.
- [4] 遠藤ほか: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).