

# 042021 用言意味属性を用いた適合文型パターンの絞り込み

計算機工学講座 C 前田 春奈

## 1 はじめに

日英機械翻訳において、用例翻訳、統計翻訳など様々な翻訳方式が提案されている。提案されている多くの翻訳方式の1つにパターン翻訳がある。パターン翻訳は日本語と英語のパターンを用いて翻訳を行う翻訳方式である。

従来、パターン翻訳を行う際、使用するパターン数が少なかったが、近年大量の文型パターンを持つ文型パターン辞書が構築された [1]。また、池原等により、日本語入力文 12 万文に対して単語レベル文型パターンで 60%以上の適合率が示されている。しかし、入力文の翻訳に使用できない文型パターンが正解パターンと共に多く抽出されてしまい、正解率が下がってしまう問題がある。

そこで、本研究では用言意味属性を利用して文型パターンの絞り込みを行い、使用する用言意味属性の条件の違いによる適合率および正解率の変化を調査する。また、調査結果より考察を行い、適合率を低下させず正解率を上げる為に最適な用言意味属性の条件を明確にする。

## 2 用言意味属性を用いた適合文型パターンの絞り込み

### 2.1 利用するデータベースと文型パターンとプログラムと用言意味属性

実験に利用するデータベースと文型パターンとプログラムと用言意味属性を以下に示す。

#### (1) 評価用試験文

実験用試験文は、日英対訳データベース (12 万文 [2]) よりランダムに選択した 110 文を使用する。

#### (2) 文型パターンデータベース

本研究では文型パターン辞書に収録された単語レベルの文型パターン (12 万対) を使用する。

#### (3) 文型パターンパーサ

文型パターンパーサ [3] は入力文 (日本語) と 12 万の文型パターンを ATN [4] を用いて照合し、入力文に適合する文型パターンを出力する。

#### (4) 用言意味属性

本研究では、池原用言意味属性体系の用言意味属性 [5] (以下、IK 用言意味属性) を使用する。

### 2.2 実験方法

実験の手順を以下に示す。

#### < 1 > 入力文と模範訳

入力文と模範訳と入力文の動詞部分に対応する用言意味属性を用意する。括弧内は用言意味属性を示す。

- 入力文 = 道路を横断する (通過迂回) ときは車に注意し (叱咤非難) なさい。
- 模範訳 = Watch out for the traffic when you cross the street.

< 2 > 文型パターンパーサによる文型パターンの抽出  
文型パターンパーサに入力文を入力し、文型パター

ンを抽出する。

#### ● <文型パターン例 1>

(抽出された日本語パターン)

/y </tk N1(は> /tcfk N2 を /cf V3(音声発話)^rentai ! ときは /tcfk N4 に /cf (V5(叱咤非難)^meirei|V5.meireigo|ND5 をしなさい)。

(英語パターン)

(V5|ND5) to N4 when <you|N1> V3 N2.

#### ● <文型パターン例 2>

(抽出された日本語パターン)

/y </tk N1 は> /tcfk N2 を /cf V3(通過迂回)^rentai ! ときは /cf (V4(叱咤非難)^meirei|V4.meireigo|ND4 をしなさい)。

(英語パターン)

Be careful when <you|N1> V3 N2.

### < 3 > 用言意味属性を用いた文型パターンの絞り込み

入力文の動詞部分の用言意味属性と抽出された文型パターンにおける日本語パターンの動詞部分の用言意味属性を比較する。もし意味属性が一致しない場合は削除する。

#### ● <削除された文型パターン (文型パターン例 1)>

(日本語パターン)

/y </tk N1(は> /tcfk N2 を /cf V3(音声発話)^rentai ! ときは /tcfk N4 に /cf (V5(叱咤非難)^meirei|V5.meireigo|ND5 をしなさい)。

(英語パターン)

(V5|ND5) to N4 when <you|N1> V3 N2.

### < 4 > 英語パターンを使った出力英文の作成

手順 3 の絞り込みで残った文型パターンにおける英語パターンの変数部分に、入力文に合った英単語を人手で代入し、出力英文とする。

#### ● <残った文型パターン (文型パターン例 2)>

(日本語パターン)

/y </tk N1 は> /tcfk N2 を /cf V3(通過迂回)^rentai ! ときは /cf (V4(叱咤非難)^meirei|V4.meireigo|ND4 をしなさい)。

(英語パターン)

Be careful when <you|N1> V3 N2.

(出力英文)

Be careful when you cross the street.

### < 5 > 出力英文の評価

すべての出力英文に対して A から C の 3 段階で評価を行う (2.3 節の評価方法参照)。

#### ● (評価)

Be careful when you cross the street.(A 評価)

### < 6 > 適合率および正解率の算出

すべての出力英文に対して評価を行った後、評価結果から適合率および正解率を算出する。

### 2.3 評価方法

出力英文の評価は人手によって行う。判断基準は以下の 3 段階とする。

- A...英語パターンの変数部分に英単語を代入すれば理想的な訳を作成可能

- B...パタ - ンの問題 (前置詞, 冠詞やパタ - ン表記ミス) を解決すれば理想的な訳を作成可能
- C...入力文に合った訳が作成不可能

## 2.4 適合率および正解率の算出方法

適合率 (以下,  $R1$ ) と正解率 (以下,  $P1$  および  $P2$ ) は以下の式でそれぞれ算出する.

- $R1 = \frac{\text{自己パターン以外のパターンが 1 つ以上抽出された文数}}{\text{入力文数}}$
- $P1 = \frac{A, B \text{ と評価されたパターン}}{\text{抽出された全パターン}}$
- $P2 = \frac{1 \text{ つ以上 } A, B \text{ と評価されたパターンを持っている文数}}{\text{自己パターン以外のパターンが 1 つ以上抽出された文数}}$

## 2.5 実験条件

用言意味属性を用いた絞り込み実験の実験条件は以下の 5 つの条件でそれぞれ行う.

1. 文法情報
2. IK 第 1 分類 (11 分類)
3. IK 第 2 分類 (66 分類)
4. IK 第 3 分類 (248 分類)
5. IK 第 4 分類 (371 分類)

## 3 実験結果

文型パターンパーサを用いて入力文 110 文と文型パターンデータベースを照合した所, 62 文が文型パターンを 1 つ以上抽出できた. 入力文は文型パターンデータベースには含まれていない. したがって, 本研究はオープンテストとなる. 絞り込み実験の結果を表 1 に示す.

表 1: 実験結果

条件	R1	P1	P2	$R1 \times P1$	$R1 \times P2$
文法情報	53% (55/103)	14% (147/1047)	49% (27/55)	7.5%	26%
IK 第 1 分類 (11 分類)	44% (45/103)	19% (53/287)	51% (23/45)	8.1%	22%
IK 第 2 分類 (66 分類)	28% (29/103)	34% (33/97)	41% (12/29)	9.6%	12%
IK 第 3 分類 (248 分類)	20% (21/103)	57% (27/47)	43% (9/21)	12%	8.7%
IK 第 4 分類 (371 分類)	19% (19/103)	58% (26/45)	42% (8/19)	11%	7.8%

(括弧内の数字は,  $R1$  および  $P2$  は文数を表し,  $P1$  はパターン数をそれぞれ表す.)

表 1 より,  $R1$ ,  $P1$  および  $P2$  において最も高い値はそれぞれ,  $R1$  は 53%(文法情報),  $P1$  は 58%(IK 第 4 分類),  $P2$  は 51%(IK 第 1 分類) であった.

## 4 考察

### 4.1 意味属性の条件による実験結果の比較

情報検索において, 一般的には  $R1 \times P1$  の値が高くなる条件を使用する. 実験結果より,  $R1 \times P1$  の値は IK 用言意味属性の第 3 分類が最も高い. しかし, IK 用言意味属性の第 3 分類を使用した場合, 残った文型パターンの総数は 47 パターンとなり, 多くの文型パターンが絞り込みで削除されてしまった. 実際の翻訳では, さらに情報を付加して最終的に翻訳に最も適した文型パターンを選択する. したがって, 意味属性を使用した文型パターンの絞り込みの過程では, 出来るだけ翻訳に使用できる文型パターンを残す事が望ましい. よって, 用言意味属性を用いた文型パターンの絞り込みでは, IK 用言意味属性の第 1 分類を使用するのが最も適していると考えている.

### 4.2 字面と用言意味属性による意味制約

実験で A 評価および B 評価と判定されたパターンは合計 147 パターンあり, 147 パターン中, パターンの動詞部分が字面で表記されているパターンは 16 パターンあった. 動詞部分が字面で表記されているパターン例を以下に示す.

- <動詞部分が字面で表記されているパターン例>  
(パターン (日))  
/y \$1^{ }{/tcfk N1 は } /tefk N2 を /cf 頼って \$1 /yef 上京した.

この 16 パターンはパターンの動詞部分に用言意味属性が付与されていないが, 字面での表記を用言の意味的制約として考えると最も厳しい意味的制約であると考えられる. よって, 実験ではパターンの動詞部分が字面で表記されている 16 パターンを, 使用する用言意味属性の条件に関係なく  $P1$  および  $P2$  の結果に含めた.

動詞部分が字面で表記されている 16 パターンの評価を調べてみた所, A と評価されたパターンが 3 パターン, B と評価されたパターンが 13 パターンであった. また, 入力文と同じ字面を含んだパターンを使用した場合の正解率は 100%であった. IK 用言意味属性の第 1 分類を使用した絞り込み実験では正解率が 51%であった事から, 字面による意味制約は用言意味属性よりも効果的であると言える.

## 5 おわりに

本研究では, 用言意味属性を用いた文型パターンの絞り込みを行い, 正解率を上げる為の最適な条件を明確なことを目的とした. 実験結果より, 単語レベルに対しての用言意味属性を用いた文型パターンの絞り込みでは, 文法情報よりも IK 第 1 分類を用いた方が正解率は高くなった.

今後はさらに適合率および正解率の値を向上させる方法を考えていく. 具体的には, 句レベルおよび節レベルの文型パターンを使用した用言意味属性による文型パターンの絞り込みを行い, 全レベルの文型パターンを用いた場合の適合率および正解率を求める予定である.

## 参考文献

- [1] 池原ほか:非線型な重文複文の表現に対する文型パターン辞書の開発, 情報処理学会, 自然言語処理研究会 2005-NL-170, pp.157-164, 2005-11.
- [2] 村上ほか:日本語英語の文対応の対訳データベース, 「言語・認識・表現」, 第 7 回年次研究会, 2002-12.
- [3] 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, 2004.
- [4] 長尾ほか:「自然言語処理」, ISBN4-00-010355-5, 岩波書店, 1996.
- [5] 池原悟:日本語単文の意味分類体系の検討, 「言語・認識・表現」研究会, 第 15 回辞書プロジェクト, 2004-9.