

## 日英文型パターンの意味的対応方式

斎藤健太郎 池原悟 村上仁一

鳥取大学大学院工学研究科

{ksaito,ikehara,murakami}@ike.tottori-u.ac.jp

### 1. はじめに

柔軟で意味的に適切な翻訳を実現する方法として、意味としてまとまりをもつ原言語<sup>1</sup>の表現をパターンとし、意味類型<sup>2</sup>を介して目的言語<sup>3</sup>のパターンと対応づける方法 [2] が注目されている。この手法は従来の単一的な対応でなく、複数対複数で対応づけられるため、原文に応じて訳語に用いるパターンを選択できるという利点がある。しかし意味類型を用いた翻訳を実現するためには、複数の候補の中から目的言語のパターンを原文に応じて一意に決定する手法が必要となる [3]。

そこで本研究はまず比較文型を例としてとりあげ、意味類型を用いて原言語のパターンに意味的・構造的に対応する目的言語のパターンを求める仕組みを示す。また訳語の表現を一意に決定できない問題点を解決するために、パターンに含まれない単語やパターン内の変数に着目し、複数の候補の中から原文に応じて訳語のパターンを一意に絞り込む手法を示す。

### 2. 従来の研究と意味類型の必要性

現在、機械翻訳では、意味としてまとまりをもつ表現をパターンとして捉え、対応づける手法が注目されている。

従来の対応づけ [1] は、動詞と名詞の組合わせでパターンが1対1で対応づけられている。そのため、対応が単一的で、原文に応じて適切なパターンを選ぶことができないという問題点がある。

その問題点を解決する方法として、意味類型を用いたパターンの対応づけが考えられる。図1に意味類型を介して日本語と英語のパターンを対応づける方法の例を示す。

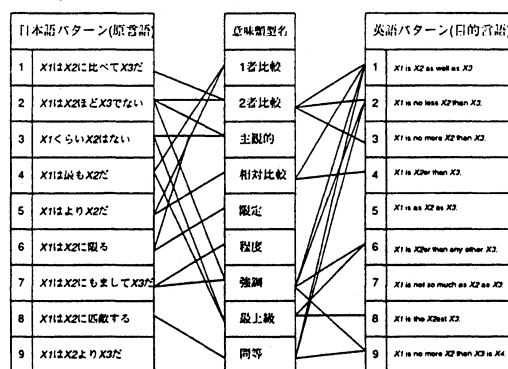


図1: 意味類型を用いた日英パターンの対応例

この手法は意味類型を介し、原言語のパターンと目的言語のパターンが複数対複数で対応されるため、原文に最適な目的言語の表現を選択できる可能性がある。しかしこの手法には、複数の目的言語のパターンの中から、原文に応じてパターンを一意に絞り込む手法が必要となる。

そこで本研究はまず比較文型を例としてとりあげ、意味類型を用いて原言語のパターンに意味的・構造的に対応する複数の目的言語のパターンを求める仕組みを示す。また訳語の表現を一意に絞り込むために、原文中のパターンに含まれない単語や変数に着目し、複数の候補の目的言語パターンの中から原文に応じて一意に絞り込む手法を示す。

### 3 パターンの対応づけ

#### 3.1 パターンの作成

本研究では、まず高校英語の参考書から比較の用例500文程度を収集し、用例を基にパターンを作成する。作成したパターンに番号をつけた表の一部を例として表1及び2に示す。

表中の「同値番」は意味と品詞が同じパターンの集合(以下、同値パターン群)ごとの番号を表し、「パ番」はパターンごとに付加した番号を示す。

<sup>1</sup>言語 A から言語 B へ翻訳する際、基となる言語 A

<sup>2</sup>意味に用いる概念を対応づける表現形式

<sup>3</sup>言語 A から言語 B へ翻訳する際、対象となる言語 B

表1：日本語のパターン表の一部

同値番	パ番	パターン
1	1A	N1はN2と同じようにKだ
	1B	N1はN2くらいKだ
2	2A	N1はN2と同じ程度FVする
3	3A	N1とN2は同じだ
	3B	N1はN2も同然だ
4	4A	N1はN2ほどKでない
	4B	N2はN1に比べてKだ
5	4C	N2はN1よりKだ(※)
	5A	N1はN2ほどFVしない
6	6A	N1はN2よりN3が劣る
	6B	N1はN2よりN3がよくない

表2：英語のパターン表の一部

同値番	パ番	パターン
1	1a	N1 is as K as N2.(※)
	1b	N1 is no less K than N2.
2	2a	N1 V as F as N2.
	2b	N1 V no less F than N2.
3	3a	N1 is the same as N2.
	3b	N1 is as good as N2.
	3c	N1 is N2 in effect.
4	4a	N1 is inferior to N2 in N3.
	4b	N1 is worse than N2 in N3.
5	5a	N1 excel over N2.
	5b	N1 is superior to N2 in N3.
6	6a	N2 is not K enough for N1.
	6b	N1 is more K than N2.(※)

(※)4Cと6bのパターンの基となった用例の対訳文を例として示す。

(例文1)Peace is more important than gold (is).

(例文2)「平和は金より大切だ」

### 3.2 意味的・構造的なパターンの対応

以上の作成したパターンを意味類型を介して1対複数の関係で対応づけるために、パターンを意味・品詞の面から対応づける仕組みが必要となる。本研究で意味と品詞で対応づける際に着目する項目を以下に示す。

1. 意味ラベル…[比較級], [同級], [最上級]のように比較の代表的な意味を表す。意味ラベルの対応により意味のおおまかな検索が可能。
2. 付加意味…[強調], [否定], [重々しい]のように比較の付加的な意味を表す。付加意味の対応により意味の詳細な検索が可能。
3. 品詞列…[N(名詞), V(動詞), K(形容詞), F(副詞)]などのパターンに用いている品詞の種類・数を表す。品詞列の対応により品詞ごとの検索が可能。

意味の概念の付加は、著者の判断に基づく。原言語パターンの意味ラベル、付加意味、品詞列を調査し、それに対応する目的言語パターンを抽出する。

パターンを対応づける手順を図2に示す。

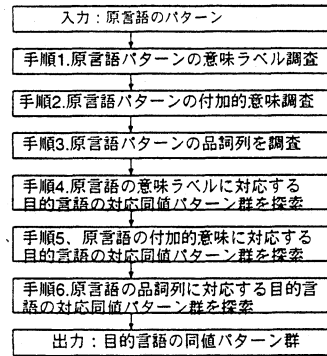


図2：パターンを対応づける仕組み

### 3.3 1対複数の関係で対応づける例

以下に意味ラベル、付加意味、品詞列を用いた対応づけの例を示す。

(例1)日本語パターン2A「N1はN2と同じ程度FVする」に対応する英語パターンを表2より探す手順を以下に示す。

手順1-3 2Aの意味ラベル=[同級], 付加意味=なし, 品詞列=[N1,N2,F,V]

手順4 意味ラベル[同級]の英語同値番号=1,2

手順5 省略

手順6 手順1の結果の中で、品詞列が[N1,N2,F,V]の英語同値番号=2

日本語パターン2Aに対応する同値パターン番号は2で、対応するパターンは2a「N1 V as F as N2.」と2b「N1 V no less F than N2.」の2つである。

例1のように対応の仕組みによって、意味と品詞が対応している同値パターン群を抽出する。しかし例1のように複数のパターンが対応するため、その中から原文に応じてパターンを絞り込む方法が必要となる。

### 4. パターンを原文に応じて絞り込む方法

原言語のパターンと目的言語パターンの関係が1対複数であるため、原文に応じて一意に目的言語のパターンを絞り込む手法が必要である。以下に原文に応じて一意に絞り込む手法を示す。

1. 原文中にパターンに含まれない単語があれば、その単語の意味(以下、パターン外意味)を利用してパターンを絞り込む。
2. 原文のすべての単語がパターンに含まれる場合、原文中の変数の意味(以下、変数意味情報)を利用して、パターンを絞り込む。
3. 上記の手法で一意に決定しない場合、あらかじめ登録しておいたパターン(以下、デフォルトパターン)を用いる。

原文に応じてパターンを一意に絞り込む手法の流れを以下の図5に示す。

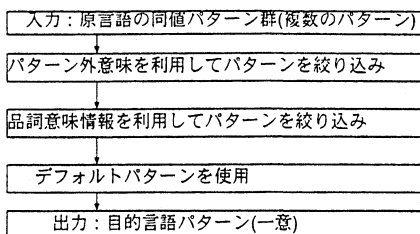


図3：パターンを一意に絞り込む手法の流れ

#### 4.1 パターン外意味による絞り込み

本研究では、パターン外意味とパターンのもつ意味を比較し、最も一致したパターンを選択する。

以下にパターン外意味を利用してパターンを絞り込む方式の例を示す。

(例文1)「私の考えも結局のところ君と同じだ」

例文1は、表1より3A「N1はN2と同じだ」にあてはまる。3Aに対応するパターンとして、表2より3パターンが出力される。出力されたパターンを意味とともに以下に示す。

3a N1 is the same as N2.[-]

3b N1 is as good as N2.[重々しい]

3c N1 is N2 in effect.[強調, 重々しい]

例文1の3Aに含まれない単語「結局のところ」には、[強調, 重々しい]の意味があると考えられる。そこで本研究の手法では、3a,3b,3cのうち最も意味が一致した3cを目的言語パターンとして出力する。

#### 4.2 変数意味情報による絞り込み

パターンに含まれない単語がない場合、パターン外意味によってパターンを絞り込むことは、不可能である。このため、本研究では、原文中のすべての単語がパターンに含まれる場合、パターンの変数のもつ意味と目的言語のパターンの変数意味情報が最も一致したパターンを選択する。

以下に変数意味情報を用いた手法の例を示す。

(例文2)「父の命令は法律同然であった」

例文2は、表1より3B「N1はN2と同然である」にあてはまる。3Bに対応するパターンとして、対応の仕組みによって表2から3パターンが出力される。出力されたパターンを変数意味情報とともに以下に示す。

3a N1 is the same as N2.[-]

3b N1 is as good as N2.[N1重々し,N2重々し]

3c N1 is N2 in effect.[N2重々しい]

例文2には、原文中のすべての単語がパターンに含まれるため、パターン外の意味を利用してパターンを絞り込むことができない。しかし例文2の「N1:父の命令」「N2:法律」の意味が[N1重々しい,N2重々しい]をもつと考えられる。そこで本研究では、上記の3a,3b,3cのうち、変数意味情報が最も一致した3bを目的言語パターンとして出力する。

#### 4.3 デフォルトパターンの設定

4.1と4.2で示した手法を用いて一意にパターンを絞り込めなかった場合、あらかじめパターンごとに登録しておいたデフォルトパターンを選択する。以下にデフォルトパターンの設定方法を示す。

1. 用例中で原言語のパターンに対応している頻度の最も高い目的言語のパターンをデフォルトパターンとして設定
2. 上記の設定方法において、用例が同数の場合や用例の存在しないパターンの場合、人手で設定

#### 5 評価実験

本研究の性能評価のため、評価実験を行う。以下の2つの調査項目を目的として行う。

1. 本研究で参考書より作成したパターン総数の調査
2. 原文より一意に絞り込む手法の性能調査

##### 5.1 パターンの総数

高校参考書の比較文500文を基にパターンを作成したパターンの総数を調査する。パターンの総数を表3に示す。

表3：パターンの総数

	日本語	英語
パターン総数	361	252

##### 5.2 原文に応じて絞り込めた割合の調査

本手法の性能を調査するため、新聞記事、アンカー英和辞典・和英辞典・学研英和辞典・基本語用例辞典・英語表現辞典の対訳コーパスから比較文を192文抽出し、本研究を適用する。

評価実験の評価方法を以下に示す。

1. カバー率は、原文より抽出した原言語パターンがパターン表に適合した割合と定義する。
2. 「平均対応パターン数」は、意味と品詞が対応するパターンの平均数と定義する。
3. 「一意対応割合」は、原言語のパターンと目的言語のパターンが1対1で対応していた割合と定義する。
4. 「原文に応じてパターン絞り込めた割合」は、パターン外意味・変数意味情報を利用して原文に応じてパターンを一意に絞り込めた割合と定義する。

表4：評価実験の結果

	英→日	日→英
1. カバー率	68 % (129/192)	74 % (141/192)
2. 平均対応パターン数	5.1	4.5
3. 一意対応割合	0 % (0/129)	8 % (14/141)
4. 原文に応じて絞り込めた割合	51 % (66/129)	59 % (84/141)

表4の2と4より一つの原言語パターンに平均5つの目的言語パターンが対応し、90%以上の文でパターンが一意に決定しないことがわかる。これによりパターンを一意に絞り込む手法の重要性が確認できる。また本手法のパターンを原文に応じて一意に絞り込む手法を英語から日本語51%は、日本語から英語は59%の割合で適用できた。

### 5.3 原文に応じて絞り込んだ結果とデフォルトパターンとの比較結果

原文に応じてパターンを一意に絞り込む手法とデフォルトパターンを用いる手法のどちらが原文に適しているかを調べるため、それぞれの結果を比較する。その結果を表5に示す。

表5：原文に応じて絞り込みとデフォルトの比較

原文に応じて絞り込んだパターンが通っていた割合	70%
デフォルトパターンが通っていた割合	30%

表5より原文に応じてパターンを絞り込んだ方が適していた割合が高い。これにより意味類型を用いて最適なパターンを原文に応じて選択できる可能性がある。しかし本研究では、最適なパターンの適否は著者の判断に基づいており、今後はネイティブによるさらに正確な判断が必要となる。

## 6 考察

### 6.1 デフォルトパターンの方が適切だった例

表5より原文に応じて絞り込んだ場合と初めからデフォルトパターンを用いた場合を比較した結果、原文に応じて絞り込んだ方が適している割合が高いことがわかった。しかしデフォルトパターンを用いた方が適していた場合もあった。その例を以下に対訳文とともに示す。

(例文3)「昨日より今日の方がやや暖かい」

例文3の対訳文：「Today it's slightly warmer than yesterday.」

例文3の原文は、4C「N1はN2よりKだ」にあてはまる。パターン外意味として、「やや」が「やわらげ」と考えられ、以下のパターンが対応する。

- パターン外意味より6a「N2 is not K enough for N1.[やわらげ, 否定]」
- デフォルトパターンとして6b「N1 is more K than N2.」

例文4に対してデフォルトパターンの方が適切と考えられる。この理由として、パターン外意味とパターンのもつ意味に一致していない部分があるため、意味的に適切でないパターンが出力されたことが挙げられる。このような場合、付加意味を重みづけし、異なる意味を選んだ場合の得点を差し引くことで、異なる意味をもつパターンを出力しない方法が考えられる。

### 6.2 パターンの作成方法の問題点

本研究では、用例を基にパターンを作成し、副詞を考慮していない。これに対して副詞を考慮してパターンを作成する方法が考えられる。以下にそれぞれのパターンの作成方法についての例を示す。

(例文4)「彼は蜂のように忙しい」

例文4を基に作成できるパターンを以下に示す。

- 「N1はN2のようにKだ」(副詞を考慮しない場合)
- 「N1はまるでN2のようにKだ」(副詞を考慮する場合)

しかし、副詞は非常に数が多く、全ての副詞をパターンに反映することは、困難と考えられる。そこで今後、上記の2つの作成方法について調査する必要がある。

## 7. おわりに

本稿では、比較文を例として取り上げ、原言語のパターンに意味的・構造的に対応する複数の目的言語のパターンを抽出する仕組みを作成した。またパターンに含まれない単語や変数の意味に着目し、複数のパターンの候補の中から原文に応じてパターンを一意に絞り込む手法を提案した。

本手法の性能調査のため、対訳コーパスを用いて評価実験を行った。その結果、英語から日本語は51%、日本語から英語は59%の割合で、パターンを原文に応じて絞り込めることがわかった。またデフォルトパターンとの比較を行った結果、原文に応じて絞り込んだパターンの方が原文に適している割合が高いことがわかった。これにより意味類型を用いて最適なパターンを選択できる可能性が見えたと言える。しかしまだ著者の判断による部分が多く、最適なパターンの適否など曖昧な部分がある。

今後は、パターンの作成方法について調査する必要がある。またネイティブの判断による最適なパターンを選択する手法の正確な精度を再調査する必要がある。

### 参考文献

- [1] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林: 日本語彙体系, 岩波書店 (1997)
- [2] 池原 悟: 意味類型に着目した日英意味辞書の研究企画, 辞書プロジェクト会議 (2000/4/1-2)
- [3] 斉藤健太郎, 池原 悟, 村上仁一: 日英比較表現のパターン化と意味的対応関係, 第7回言語処理学会, pp.241-244 (2001.3)
- [4] グループ・ジャマシイ: 日本語文型辞典, くろしお出版 (1998)
- [5] 斉藤 健太郎, 池原 悟, 村上 仁一: 意味類型を用いた日英文型パターンの対応づけ, 第147回自然言語処理研究会, pp.1-8 (2002.1)