

概要

近年，日英機械翻訳方式として，文型パターンを用いて翻訳を行うパターン翻訳が注目されている．パターン翻訳は，日本語の入力文に合った日本語パターンを文型パターンデータベースから検索し，検索し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生成する翻訳方式である．パターン翻訳を行う際，大量の文型パターンを用意する必要がある．過去の研究では，パターン作成における変数化の問題などがあり，意味的に独立したパターンを大量に作成する事が困難であった．これに対して [1] で提案した方法により，大量のパターンを作る事が可能になった．現在，日本語の重複文 12 万文に対してパターンが作成されている．また，池原等 [2] により，日本語入力文 12 万文に対して 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，用言意味属性を使用して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した．

具体的な調査方法として，まず文型パターンパーサを利用して文型パターンを抽出する．次に，抽出された文型パターンの日本語パターンを用言意味属性を用いてパターンの絞り込みを行う．そして，残ったパターンに対して人手で評価を行う．最後に適合率および正解率を算出する．

調査の結果，入力文 110 文中，実験 1(jpp_20040816, 3LevelPattern_ver5.3.1) では 38 文，実験 2(jpp_20050627, 3LevelPattern_ver9.0.0) では 62 文が 1 つ以上の日本語パターンの候補を抽出できた．次に，用言意味属性を用いたパターンの絞り込みを行った後，正解率を算出した所，実験 1 では用言意味属性を使用しない場合の正解率は 12%，用言意味属性 (IK 第 1 分類) を使用した場合の正解率は 17%であった．また，実験 2 では用言意味属性を使用しない場合の正解率は 14%，用言意味属性 (IK 第 1 分類) を使用した場合の正解率は 18%であった．

よって，本研究では IK を用いた文型パターンの絞り込みによって，正解率が向上した事を確認した．

目次

1	はじめに	1
2	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	2
2.1	評価用試験文	2
2.2	文型パターンデータベース	3
2.2.1	パターン記述言語	5
2.3	文型パターンパーサ	7
2.4	用言意味属性	10
3	用言意味属性を用いた適合文型パターンの絞り込み	11
3.1	実験方法	11
3.2	評価方法	14
3.3	適合率および正解率の算出方法	19
3.4	ヒット率	20
3.5	出力英文作成時の注意点	21
3.5.1	日本語パターンの注意点	21
3.5.2	英語パターンの注意点	21
4	実験結果	23
4.1	実験条件	23
4.2	実験 1(jpp_20040816, 3LevelPattern_ver5.3.1)	25
4.3	実験 2(jpp_20050627, 3LevelPattern_ver9.0.0)	27
5	考察	29
5.1	意味属性の条件による実験結果の比較	29
5.2	IK と NY の結果比較	29
5.3	字面と用言意味属性による意味制約	30
5.4	名詞意味属性の利用	31
5.5	句レベルおよび節レベルパターンの利用	32
5.6	意味属性の付与誤り	33
5.7	各絞り込み条件で消えた例	34

目 次

1	IK の例	10
2	$R1$ の結果 (実験 1)	26
3	$P1$ の結果 (実験 1)	26
4	$P1$ の結果 (実験 1)	26
5	$R1$ の結果 (実験 2)	28
6	$P1$ の結果 (実験 2)	28
7	$P2$ の結果 (実験 2)	28
8	名詞意味属性の例	31

表 目 次

1	変数一覧	5
2	関数一覧	6
3	記号一覧	6
4	離散記号の適合条件	6
5	NY の例	10
6	IK 第 1 分類	23
7	IK 第 2 分類	23
8	IK 第 3 分類	24
9	IK 第 4 分類	24
10	NY	24
11	実験 1 結果	25
12	実験 2 結果	27

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、大量の文型パターンを用いて翻訳を行う翻訳方式である [1]。従来、パターン翻訳を行う際、使用する文型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 [2]。また、[2]により、日本語入力文 12 万文に対して 60%以上の再現率が示されている。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。具体的な調査方法として、まず文型パターンパーサを利用して文型パターンを抽出する。次に、抽出された文型パターンの日本語パターンを用言意味属性を用いてパターンの絞り込みを行う。そして、残ったパターンに対して人手で評価を行う。最後に適合率および正解率を算出する。

調査の結果、入力文 110 文中、実験 1(jpp_20040816, 3LevelPattern_ver5.3.1) では 38 文、実験 2(jpp_20050627, 3LevelPattern_ver9.0.0) では 62 文が 1 つ以上の日本語パターンの候補を抽出できた。次に、用言意味属性を用いたパターンの絞り込みを行った後、正解率を算出した所、実験 1 では用言意味属性を使用しない場合の正解率は 12%、用言意味属性 (IK 第 1 分類) を使用した場合の正解率は 17%であった。また、実験 2 では用言意味属性を使用しない場合の正解率は 14%、用言意味属性 (IK 第 1 分類) を使用した場合の正解率は 18%であった。よって、本研究では IK 用いた文型パターンの絞り込みによって、正解率が向上した事を確認した。

本稿は以下の章で構成されている。2 章では、実験で使用するデータベースとパターンとプログラムと用言意味属性について説明する。3 章では用言意味属性を用いたパターンの絞り込み実験の実験手順について説明する。4 章では実験結果を示す。5 章では、実験結果より考察を行い、最後に 6 章で本研究の結論を述べる。

2 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

2.1 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 [3]) よりランダムに選択した 110 文を使用する。例文を以下に示す。また、付録 1 に評価様試験文 110 文と模範訳を載せる。

- <評価用試験文例 1>

(入力文)

彼にはその任務を果たせるだけの能力がなかった。

(模範訳 1)

He was not equal to the task.

(模範訳 2)

He was not up to the job.

(模範訳 3)

The man was not a match for the duty.

- <評価用試験文例 2>

(入力文)

彼は 10 歳になるまでそこに住んでいた。

(模範訳 1)

He lived there until he was ten.

(模範訳 2)

He resided there up until he was ten.

(模範訳 3)

He inhabited that place until he was ten years.

- <評価用試験文例 3>

(入力文)

日本の人口はイギリスとフランスを合わせたより多い。

(模範訳 1)

Japan's population is larger than that of Britain and France combined.

(模範訳 2)

Japans's inhabitants number more than that of Britain and France put together.

(模範訳 3)

Japan's population is greater than that of Britain and France added.

評価用試験文例 1, 評価用試験文 2 および評価用試験文 3 に示されるように, 文型パターンパーサに入力される入力文 1 文に対して模範訳は 3 つ用意されている。

2.2 文型パターンデータベース

文型パターンは, 日英言語の表現対において, 線形要素を変数記号に書き換えたパターンである [1]. 本研究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. この辞書には, 単語レベルのほか, 句レベル, 節レベルが収録されており, その中の単語レベルは表現に含まれる名詞, 動詞, 形容詞, 副詞などの自立語の線形な要素を変数化している. 単語レベル文型パターンの例を以下に示す. 単語レベル文型パターンは, 日英文型パターン対作成時に使用した対訳 (原文) と原文を元に作成された日英文型パターン対が収録されている. 以下に例を示す.

- <単語レベル文型パターンデータベース例 1>

(日本語原文)

濃霧が発生して忽ちにして視界をさえぎった。

(英語原文)

The field of vision was obstructed immediately after a fog came out.

(日本語パターン)

WJAO003535-00: /ytcfk N1 が /cfV2(NY:15,NY:26,NY:26,IY:5310)(て|で)#1[/y ADV4] /tck N5 を /cfV6(NY:05,NY:16,NY:20,IY:3410,IY:5220,IY:5950,IY:8330).kako。

(英語パターン)

WEAO003535-00:N5 V6.past.passive #1[ADV4] after N1 V2.past.

単語レベル文型パターンデータベース例 1 より, 日本語パターンの「*WJAO003535-00*」は単語レベル日本語文型パターンのパターン ID を表し, 「*ADV*」は副詞を表し, 「*.kako*」

は過去形を表している。また、*NY* は日本語語彙大系の用言意味属性番号を表し、*IY* は池原用言意味属性の意味属性番号を表している。

一方、英語パターンの「*WEAO003535-00*」は単語レベル英語文型パターンのパターンIDを表している。また、「*.past*」は過去形を表し、「*.passive*」は受動態を表している。

- <単語レベル文型パターンデータベース例 2>

(日本語原文)

彼のお母さんがああ若いとは思わなかった。

(英語原文)

I never expected his mother to be so young.

(日本語パターン)

WJAC000004-00: /y </tk N1 は> /tcfk N2 の /k N3 が /tcfk ああ /f AJ4(*NY*:05,*IY*:9300)

とは /yf V5(*NY*:31,*NY*:32,*IY*:0200,*IY*:0262,*IY*:1111).*hitei.kako*。

(英語パターン)

WEAC000004-00: <I|N1> never V5.past N2[^] poss N3 to be so AJ4.

同様に、単語レベル文型パターンデータベース例 2 より、日本語パターンの「*AJ*」は形容詞を表している。また、「*.hitei*」は否定形を表している。

一方、英語パターンの「[^]poss」は名詞の所有格変化を表している。

- <単語レベル文型パターンデータベース例 3>

(日本語原文)

タイプライターを捨ててワープロを選んだ。

(英語原文)

I have abandoned my typewriter in favor of a word processor.

(日本語パターン)

WJAQ000043-00: /y </tk N1 は> /tcfk N2 を /cf V3(*NY*:23,*NY*:27,*NY*:32,*IY*:0342,

IY:8423)(て | で) /ytck N4 を /cf 選んだ。

(英語パターン)

WEAQ000043-00: <I|N1> V3.pft <my|N1[^] pron[^] poss> N2 in favor of N4.

同様に, 単語レベル文型パターンデータベース例3より, 英語パターンの「.pft」は完了形を表し, 「.pron」は名詞の目的格変化を表している.

2.2.1 パターン記述言語

文型パターンを構成する要素には, 字面, 変数, 関数, 記号がある. 変数には, 名詞を表す *N*, 動詞を表す *V*, 名詞句を表す *NP* などがある (表1). また, 関数は, 変数に適合する値の形式や, 字面の指定, 表現の統括を行う (表2). 一方, 記号は, パターン要素の適合の仕方について, 任意化, 選択, 順序変更, 記憶という制御を行う (表3および表4).

表 1: 変数一覧

変数名	意味	変数名	意味
単語レベル		句レベル	
<i>N</i>	名詞	<i>NP</i>	名詞句
<i>ND</i>	「する」に先行する名詞	<i>VP</i>	動詞句
<i>NUM</i>	数詞	<i>AJP</i>	形容詞句
<i>TIME</i>	時詞	<i>AJVP</i>	形容動詞句
<i>V</i>	動詞	<i>ADVP</i>	副詞句
<i>AJ</i>	形容詞	節レベル	
<i>AJV</i>	形容動詞	<i>CL</i>	節
<i>ADV</i>	副詞		
<i>REN</i>	連体詞		
<i>GEN</i>	限定詞		

表 2: 関数一覧

種類	意味と例
様相関数	変数に後続する表現の形式を指定 例) $V1.reru$
語尾関数	変数に対応する表現の形式を指定 例) $V1^{katei}$
字面関数	変数に含まれる字面を指定 例) $\#大変(CL1)$
マクロ関数	引数内表現を指定する変数が統括 例) $\#CL1(N2がV3)$

表 3: 記号一覧

記号名	表記	意味
離散記号	/...	文型に無関係な要素
選択記号	(...)	いずれかの要素列と適合
任意記号	[...]	文型選択上, 任意の要素
補完要素記号	<...>	ゼロ代名詞等
順序任意要素 指定記号	{...,...}	順序入れ替え可能な範囲
位置変更可能 要素指定記号	$\$n^{\{...\}}, \n	指定位置に入れ替え可能
文節境界記号	!	文節の境界と適合
記憶記号	$\#n$	適合内容を記憶

表 4: 離散記号の適合条件

表記	適合条件
y	連用節
t	連体節
c	格要素
f	連用修飾句 (副詞句)
k	連体修飾句 (形容詞, 形容動詞連体形, 連体詞)

2.3 文型パターンパーサ

文型パターンパーサは [4], 日本語文と文型パターンを ATN[5] を用いて照合し, 日本語文に適合する文型パターンを出力する. 文型パターンパーサに入力文を入力すると, 入力文に適合する文型パターンを複数出力する.

本研究では, 2つのバージョンの文型パターンパーサと, 2つのバージョンの文型パターンファイルを使用した. 使用した文型パターンパーサと文型パターンファイルのバージョンを以下に示す.

1. (実験 1)

- 文型パターンパーサ: jpp_20040816
- 文型パターンファイル: 3LevelPattern_ver5.3.1

2. (実験 2)

- 文型パターンパーサ: jpp_20050627
- 文型パターンファイル: 3LevelPattern_ver9.0.0

<文型パターンパーサを用いた適合文型パターン抽出例 (実験 1) >

文型パターンパーサに日本語文を入力すると, 日本語文に適合する文型パターンを複数出力する. 以下に例を示す. マッチパターン数とは, 入力文に対して, 抽出された適合文型パターンの数である.

(入力文)

先生は生徒の学力に合わせた試験問題を作った。

(マッチパターン数)

25パターン

- <抽出された文型パターン例 1>

(日本語パターン)

*WJ246589-01: /y \$1^{/tk N1 は} /tcfk N2 に /cf (V3.kako^rentai| ND3 をした)
! N4(NI:1056,NI:1041) を \$1 /cf V5kako。*

(英語パターン)

WE202268-01: N1 V5.past N4 (V3|V(ND3)).past on N2.

- <抽出された文型パターン例 2>

(日本語パターン)

WJ155965-01: /y </tk N1 は> /tcfk N2 に /cf V3.kako^{rentai} ! N4 を /cf V5.kako。

(英語パターン)

WE144313-00: N1 V5.past N4 that had V3 on N2.

- <抽出された文型パターン例 3>

(日本語パターン)

WJ145378-01: /y </tk N1 は> /tcfk N2 に /cf V3(て | で)#4[/f AJV5^{rentai}] ! N6 を /cf V7.kako。

(英語パターン)

WE136309-00: N1 V7.past #4[AJ5] N6 that V3.past to N2.

<文型パターンパーサを用いた適合パターン抽出例 (実験 2) >

(入力文)

先生は生徒の学力に合わせた試験問題を作った。

(マッチパターン数)

31 パターン

- <抽出された文型パターン例 1>

(日本語パターン)

WJAG010519-00: /y \$1^{/tcfk N1 は} #1[/cf GEN3] /k N4 の /k N5 に /cf (V6.kako^{rentai} | ND6 をした) /f N7.kako。

(英語パターン)

WEAG010519-00: N1 V8.past #1[AJ3] N4 the AJ(V6 | ND6) N7 for N4^{pron} poss N5.

- <抽出された文型パターン例 2>

(日本語パターン)

WJBD026841-00: /y \$1^{/tcfk N1 は} /tcfk N2 に /cf (V3.kako^{rentai} | ND3 をした) /f N4 を \$1 /cf V5.kako。

(英語パターン)

WEBD026841-00: N1 V5.past N4 V(V3 | ND3).past on N2.

- <抽出された文型パターン例 3>

(日本語パターン)

WJAC007552-00: /y \$1^{/tcfk N1 は} #1{/cf V2.kako^rentai /f N3 を \$1, /tcfk N4 に} /cf V5#[.teyaru].kako。

(英語パターン)

WEAC007552-00: N1 V5.past N3 N1^pron had V2.past to N4.

2.4 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 [6] と日本語語彙大系の用言意味属性 [7] を使用する。

池原用言意味属性体系の用言意味属性 (以下, IK) は用言が 4 桁の数字を用いて 4 段階に分類されている。IK は使用する桁数により第 1 分類から第 4 分類まで意味を分ける事ができ、使用する桁数が多いほどより細かい意味が表されている。IK の例を以下に示す。なお、付録 2 に IK と所属単語を載せる。

	0	2	1	1	
第1分類(意味知覚と情緒の表現)	0				…第1分類(11分類)
第2分類(意味個人的感情)		02			…第2分類(66分類)
第3分類(意味喜怒哀)			021		…第3分類(248分類)
第4分類(意味歓喜)				0211	…第4分類(371分類)

図 1: IK の例

また、日本語語彙大系 [7] の用言意味属性 (以下, NY) は用言が 36 に分類されている。NY の例を表 5 に示す。なお、付録 3 に NY 体系を載せる。

表 5: NY の例

意味属性コード	意味分類	所属単語例
18	身体動作	上京する, 行く, 帰る
23	感情動作	起きる, 触る, 書く
31	物理的移動	感動する, 泣く, 笑う

3 用言意味属性を用いた適合文型パターンの絞り込み

3.1 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, 適合する文型パターン全てを文型パターンデータベースから抽出する.

- <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.

文型パターン例 1 および文型パターン例 2 に示されるように, 入力文を文型パターンパーサに入力すると複数の文型パターンが抽出される. 本研究では 1 入力文に対して抽出される文型パターンの最大数を 100 パターンとし, 100 パターン以上抽出された場合はヒット率 (3.4 節参照) の高い上位 100 パターンを選択する.

また、文型パターン例1に示されている変数 N は名詞を表し、変数 V は動詞を表し、変数 ND は用言性名詞を示している。さらに、 \hat{rentai} 、 \hat{meirei} および $.meireigo$ は関数を表し、関数の意味はそれぞれ、 \hat{rentai} は動詞の連体形変形である事を示し、 \hat{meirei} および $.meireigo$ は動詞に続く語尾が命令形である事を示している。

< 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)

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

(英語パターン)

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

対応している入力文の動詞部分と抽出された文型パターンの日本語パターンの動詞部分(変数 V の部分)の意味属性がすべて一致しない場合、文型パターンは削除される。上記の例(削除された文型パターン)の日本語パターンの動詞部分と入力文の動詞部分の意味属性を比較すると、「 $V5(叱咤非難)$ 」と「注意する(叱咤非難)」で一致しているが、「 $V3(音声発話)$ 」と「横断する(通過迂回)」で一致していない。よって、上記の文型パターンは削除される。

< 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)

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

(英語パターン)

$Be careful when <you|N1> V3 N2.$

具体的な出力英文作成方法は、残った文型パターンの日本語パターンに対応する英語パターンの変数部分にそれぞれ、 $V3$ に、入力文の「横断する」に対応する英単語「cross」

を代入し、N2に、入力文の「通り」に対応する英単語「the street」を代入して出力英文を作成する。出力英文を以下に示す。

(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 (3.2 節の評価方法参照)。

- (評価)

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

上記の英文は模範訳とは異なっている。しかし、英文自体は非常に質の高い英文となっているので評価 A となる。

< 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。

3.2 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。

< 1 > A...英語パターンの変数部分に単語を代入すれば理想的な訳を作成可能

- <評価 A の例 1>

(入力文)

私は子供の将来を思うと切ない。

(模範訳 1)

I get distressed when I think of my children's future.

(模範訳 2)

I get upset as I think of my kids future.

(模範訳 3)

I get distraught when I contemplate my children's prospects.

(英語パターン)

WEBD015758-00:N1 'get' #2(^present|^past) stressed out when N1 V5 #1[N1^poss N3]'s N4.

(出力英文)

I get stressed out when I think of my children's future.

評価 A の例 1 では、模範訳とは異なる英文が出力されているが、非常に質の高い英文であるので、評価 A となる。

- <評価 A の例 2>

(入力文)

体裁の好い本だ。

(模範訳 1)

The book is elegantly got up.

(模範訳 2)

The book is stylishly decorated.

(模範訳 3)

The book is pleasingly adorned.

(英語パターン)

WEAN093870-00:N1 be smartly got up.

(出力英文)

This book is smartly got up.

同様に、評価 A の例 2 でも、模範訳とは異なる英文が出力されているが、非常に質の高い英文であるので、評価 A となる。

- <評価 A の例 3>

(入力文)

城を取り巻いて攻撃した。

(模範訳 1)

They surrounded the castle and attacked.

(模範訳 2)

They encircled the stronghold and attacked.

(模範訳 3)

They surrounded the castle and assaulted it.

(英語パターン)

WEAN102095-00: <I|N1> V4.past N3 and V5.past.

(出力英文)

They surrounded the castle and attacked.

評価 A の例 3 では、模範訳 1 と同じ英文が出力英文として出力された。よって、評価 A となる。

< 2 > 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

- <評価 B の例 1>

(入力文)

私は彼を頼って上京した。

(模範訳 1)

I came to Tokyo counting on his help.

(模範訳 2)

I arrived in Tokyo relying on his help.

(模範訳 3)

I turned up to Tokyo counting on his assistance.

(英語パターン)

WEAG009528-00: <I|N3> N(V4|ND4) where <I|N1> could rely on <my|N1^pron^poss> N2.

(出力英文)

I went to Tokyo where I could rely on my he.

評価Bの例1では、出力英部の後半部分が「my he」となっている。正確には「my he」ではなく「him」である。しかし、この部分の出力英文のミス以外は質の高い英文が作成されているので評価Bとなる。

- <評価Bの例2>

(入力文)

先生は生徒の学力に合わせた試験問題を作った。

(模範訳1)

The teacher made up a test appropriate for their scholastic abilities.

(模範訳2)

The tea cher devised a test suitable for their scholastic capabilities.

(模範訳3)

The tutor made up an exam appropriate for their academic abilities.

(英語パターン)

WEBD026841-00:N1 V5.past N4 V(V3|ND3).past on N2.

(出力英文)

The teacher made up a test appropriated for the student.

評価Bの例2では、出力英文に入力文の「学力」に相当する英文「scholastic abilities」が不足している。しかし、この部分の出力英文のミス以外は質の高い英文が作成されているので評価Bとなる。

- <評価Bの例3>

(入力文)

親に隠れて煙草を吸った。

(模範訳1)

I hid from my parents and smoked a cigarette.

(模範訳2)

I secreted myself from my parents and smoked a cigarette.

(模範訳 3)

I hid out of sight of my mum and dad and smoked a cigarette.

(英語パターン)

WEAQ218660-00: <We|N1> V3.past on N2 and V5.past N4.

(出力英文)

I hid from the parents and smoked a cigarette.

評価 B の例 3 では、出力英文の「parents」の前が「my」ではなく「the」となっている。「my」を出力するためには英語パターンの V3 の前に $N1^{poss}$ が必要となる。しかし、この部分の出力英文のミス以外は質の高い英文が作成されているので評価 B となる。
< 3 > C...入力文に合った訳が作成不可能

- <評価 C の例 1>

(入力文)

母は赤ん坊をあやして笑わせた。

(模範訳 1)

Mother played with the baby and got him to smile.

(模範訳 2)

Mum played with the infant and made him to smile.

(模範訳 3)

Mother had fun with the little one and got him to beam.

(英語パターン)

WEAC043816-00:N1 V3.past N1^pron^poss N2 V4.

(出力英文)

Mother playerd her baby smile.

- <評価 C の例 2>

(入力文)

雄弁は政治家に必要な属性だ。

(模範訳 1)

Eloquence is a necessary attribute in a politician.

(模範訳 2)

Expressiveness is an indispensable attribute in a politician.

(模範訳 3)

Persuasiveness is an essential quality in a politician.

(英語パターン)

WEAF027972-00:N1 N4 be AJ3 to N2.

(出力英文)

Eloquence attribute is necessary in a politician.

- <評価 C の例 3>

(入力文)

警官が来て騒ぎを鎮めた。

(模範訳 1)

The policemen came and got things under control.

(模範訳 2)

The police officer arrived and got things in hand.

(模範訳 3)

The policemen turned up and got stuff under control.

(英語パターン)

WEAO003535-00:N5 V6.past.passive #1[ADV4] after N1 V2.past.

(出力英文)

The disturbance quieted passive after the policemen came.

評価 C の例 1 ~ 例 3 は、どれも入力文の意味を表していない。よって評価はすべて C となる。

3.3 適合率および正解率の算出方法

適合率 (以下 $R1$) と正解率をそれぞれ算出する. 正解率については以下の2通りの評価を行う.

1. 全ての入力文に対して抽出された全ての文型パターンについて, A および B と評価された文型パターンが存在する割合 (文型正解含有率 (以下 $P1$))
2. 入力文に対して A および B と評価された文型パターンが1つでも存在する割合 (適合文型正解含有率 (以下 $P2$))

$R1$, $P1$ および $P2$ のは以下の式でそれぞれ求める.

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

(例): 入力文 100 文中, 自己パターン以外のパターンを1つ以上抽出できた入力文が 50 文, 抽出された文型パターンの総数が 1000 文, A または B と評価された文型パターンが 200 文, 1つ以上 A または B と評価された文型パターンを持っている文が 10 文であった場合, $R1$, $P1$ および $P2$ はそれぞれ以下の値となる.

- $R1 = \frac{50}{100} = 0.2(20\%)$
- $P1 = \frac{200}{1000} = 0.4(40\%)$
- $P2 = \frac{10}{50} = 0.2(20\%)$

3.4 ヒット率

ヒット率とは、入力文と適合したパターンの文字単位での一致率である。ヒット率は文型パターンパーサから得られる。例を以下に示す。

(入力文)

私は子どもの将来を思うと切ない。

- <ヒット率 100%の例>

(適合した文型パターンとマッチしている部分 (下線部分))

私は子どもの将来を思うと切ない。

(日本語パターン)

$N1$ は/ $N2$ の/ $TIME3$ を/ $V4$ と/切ない。

(英語パターン)

$N1$ 'get' #2($\wedge present|\wedge past$) stressed out when $N1$ $V5$ #1[$N1$ $\wedge poss$ $N3$]'s $N4$.

(出力英文)

I get stressed out when I think of my children's future.

- <ヒット率 56%の例>

(適合した文型パターンとマッチしている部分 (下線部分))

私は子どもの将来を 思うと切ない。

(日本語パターン)

$N1$ は/ $V2$ と/ $AJ3$ 。

(英語パターン)

$N1$ 'is' #1($\wedge present|\wedge past$) $AJ3$ to $V2$ #2($\wedge present|\wedge past$) to.

(出力英文)

I is painful tp think to.

ヒット率 100%の例では、入力文に対して質の高い英文が作成できた。一方、ヒット率 56%の例では、入力文の意味に合わない英文が作成されている。よって、一般的にヒット率が高い日本語パターンを選択した方が、品質の高い出力英文を作成できる可能性が高い。

3.5 出力英文作成時の注意点

出力英文作成時には、以下に注意する必要がある。

3.5.1 日本語パターンの注意点

入力文に適合した文型パターンを使用して英文作成を行う際、ヒット率(3.4節参照)の問題により適合した文型パターンに関係なく質の高い英文を作成できない場合がある。例を以下に示す。

- <英文作成可能例>

(入力文)

私は親に隠れてごはんを食べた。

(適合した文型パターンとマッチしている部分(下線部分))

私は親に隠れて ごはんを 食べた。

(日本語パターン)

N1は / N2に / V3て / V4。

- <英文作成不可能例>

(入力文)

私は親に隠れて煙草を吸った。

(適合した文型パターンとマッチしている部分(下線部分))

私は親に隠れて 煙草を 吸った。

(日本語パターン)

N1は / N2に / V3て / V4。

英文作成可能例では、入力文と適合した日本語パターンを比較すると、入力文の「ごはんを」がマッチしていない。パターン翻訳では、入力文のマッチしていない部分は後に要素合成法で加える。「ごはん」と「食べる」は独立した関係(線形関係)であり、それぞれの単語は別々に英語辞書から検索できるため、この例は英文作成可能となる。

一方、英文作成不可能例では、入力文と適合した日本語パターンを比較すると、入力文の「煙草」がマッチしていない。動詞「吸う」を英語に翻訳する場合、「吸う」にはいくつもの訳があり、入力文の意味を示す「smoke」を出すには「吸う」対象が「煙草」でなければならない。つまり、「煙草」と「吸う」は独立していない関係(非線形関係)であり、実際の翻訳で理想的な訳を作成するためには「煙草」と「吸う」のどちらも文型パターンに適合していなければならない。

3.5.2 英語パターンの注意点

本研究では、英語パターンを使用して出力英文を作成する場合、前置詞および冠詞については任意で英語パターンに追加(または削除)した。例を以下に示す。

- <変更可能英語パターン例>

(英語パターン)

N1 V2 in N3.

変更可能英語パターン例の英語パターンを使用して英文作成を行う場合、*N1* 及び *N3* には入力文の意味を表す適切な英単語が選ばれる。名詞を英語で表す際、任意で冠詞 (*a, the* など) が必要な場合がある。冠詞が必要な場合は人手で任意に英語パターンに冠詞を追加した。また、英語パターンに表記されている前置詞「*in*」は、通常、前の動詞または後ろの名詞によって変化 (*on, at, with*, または必要無しなど) する。よって、英語パターンに表記されている前置詞は任意で変更可能とした。

一方、接続詞などの文の文法構造を決定する表記は任意で変更できない。例を以下に示す。

- <変更不可能英語パターン例>

(英語パターン)

N1 V2 and V3.

変更不可能英語パターン例では、接続詞「*and*」がパターンに字面表記されている。「*and*」は文の文法構造を決定する重要な要素である。よって、英語パターンに字面表記されている、文の文法構造を決定する要素 (*when, that, which* など) は任意に変更できない。

4 実験結果

4.1 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6つの条件でそれぞれ行う。

(条件1) 文法情報:用言意味属性を用いた文型パターンの絞り込みを行わず, 文型パターンパーサで抽出した全ての文型パターンに対して, 適合率および正解率を算出する。

(条件2) IK 第1分類 (11分類):文型パターンパーサで抽出した全ての文型パターンに対して, 池原用言意味属性の第1分類を用いて絞り込みを行い, 絞り込まれた文型パターンパターンに対して適合率および正解率を算出する。IK 第1分類の例を表6に示す。

表 6: IK 第1分類

意味属性コード	意味分類
0	知覚の情報の表現
1	知的な行為の表現
2	日常生活の行為

(条件3) IK 第2分類 (66分類):IK の第2分類を用いて絞り込みを行う。IK 第2分類の例を表7に示す。

表 7: IK 第2分類

意味属性コード	意味分類
02	個人的感情
12	言語と表現の行為
22	家庭生活の行為

(条件4) IK 第3分類 (248分類):IK の第3分類を用いて絞り込みを行う。IK 第3分類の例を表8に示す。

表 8: IK 第3分類

意味属性コード	意味分類
021	喜悲
121	音声
223	衣食住

(条件5) IK 第4分類 (371分類):IK の第4分類を用いて絞り込みを行う。IK 第4分類の例を表9に示す。

表 9: IK 第4分類

意味属性コード	意味分類
0211	歡喜
1211	音声発話
2231	住まい

(条件6) NY(36分類):日本語語彙大系で分類された用言意味属性を用いて絞り込みを行う。NYの例を表10に示す。

表 10: NY

意味属性コード	意味分類
05	属性
23	身体動作
31	感情動作

4.2 実験1(jpp_20040816, 3LevelPattern_ver5.3.1)

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、38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、実験1はオープンテストとなる。

文型パターンを1つ以上抽出できた38文の入力文中、2文は文に動詞が含まれていないため用言意味属性を用いた絞り込みの効果を確認できない。よって、動詞が含まれていない2文については対象外とし36文について絞り込み実験を行った。対象となった36文に対して、抽出できた文型パターンの総数は684パターンで、抽出できた684パターン中、動詞部分が字面で表記されているパターンが4パターンあった。

動詞部分が字面で表記されている4パターンについては、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。

絞り込み実験の結果を表11に示す。また、 $R1$ 、 $P1$ および $P2$ のグラフを次ページに示す。さらに、付録4に抽出された全ての文型パターンの評価を載せる。

表 11: 実験1 結果

条件	$R1$	$P1$	$P2$	$R1 \times P1$	$R1 \times P2$
文法情報	33% (36/108)	12% (85/684)	55% (21/38)	4.1%	18%
IK 第1分類 (11分類)	24% (26/108)	17% (25/146)	48% (14/26)	4.1%	13%
IK 第2分類 (66分類)	19% (12/108)	40% (19/48)	48% (10/21)	7.7%	9.3%
IK 第3分類 (248分類)	15% (16/108)	52% (15/29)	50% (8/16)	7.7%	7.4%
IK 第4分類 (371分類)	11% (12/108)	57% (13/23)	50% (6/12)	6.3%	5.6%
NY (36分類)	21% (23/108)	22% (31/140)	62% (13/21)	4.1%	18%

(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11より、 $R1$ 、 $P1$ および $P2$ において最も高い値はそれぞれ、 $R1$ は33%(文法情報)、 $P1$ は57%(IK第4分類)、 $P2$ は55%(文法情報)であった。また、NYを用いた場合、 $R1$ は21%、 $P1$ は22%、 $P2$ は62%であった。

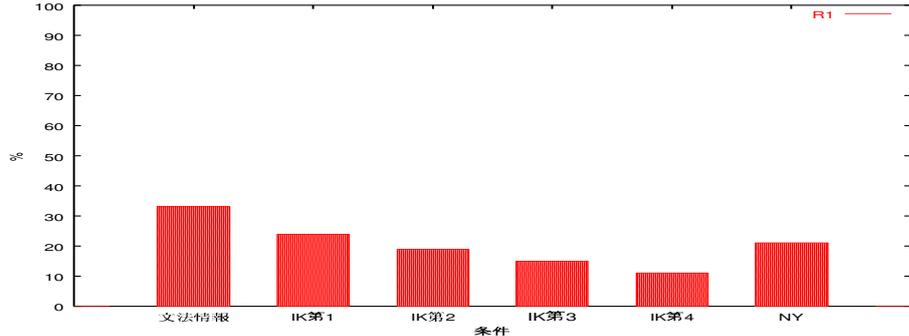


図 2: R1 の結果 (実験 1)

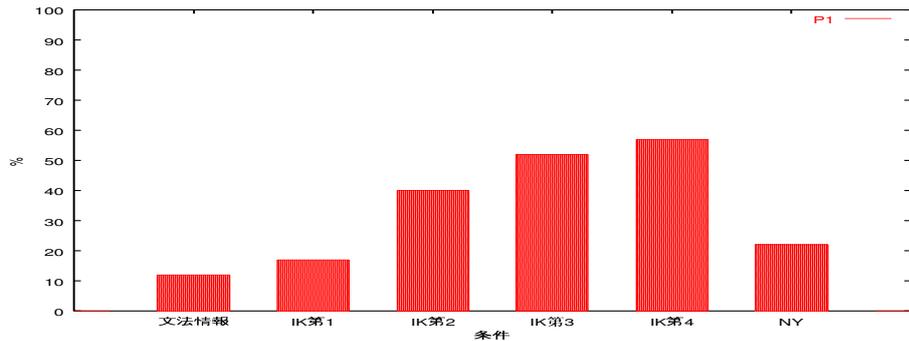


図 3: P1 の結果 (実験 1)

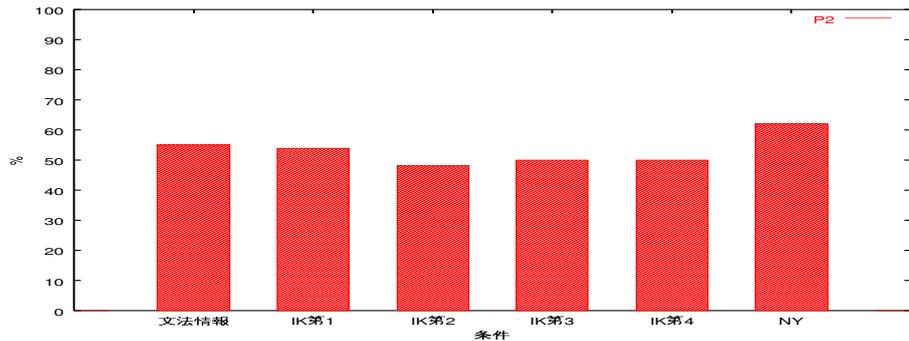


図 4: P1 の結果 (実験 1)

4.3 実験 2(jpp_20050627, 3LevelPattern_ver9.0.0)

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、実験 2 はオープンテストとなる。

文型パターンを 1 つ以上抽出できた 62 文の入力文中、7 文は文に動詞が含まれていないため用言意味属性を用いた絞り込みの効果を確認できない。よって、動詞が含まれていない 7 文については対象外とし 55 文について絞り込み実験を行った。対象となった 55 文に対して、抽出できた文型パターンの総数は 1047 パターンで、抽出できた 1047 パターン中、動詞部分が字面で表記されているパターンが 16 パターンあった。

動詞部分が字面で表記されている 16 パターンについては、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。

絞り込み実験の結果を表 12 に示す。また、 $R1$ 、 $P1$ および $P2$ のグラフを次ページに示す。さらに、付録 5 に抽出された全ての文型パターンの評価を載せる。

表 12: 実験 2 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%
NY (36 分類)	41% (42/103)	19% (66/341)	50% (21/42)	7.9%	20%

(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 12 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。また、NY を用いた場合、 $R1$ は 41%、 $P1$ は 19%、 $P2$ は 50%であった。

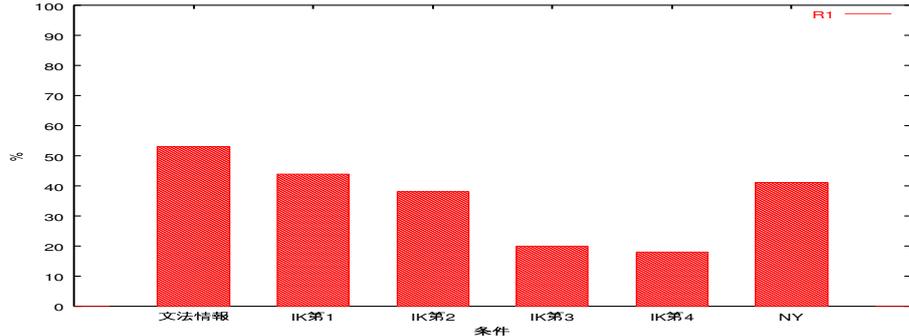


図 5: R1 の結果 (実験 2)

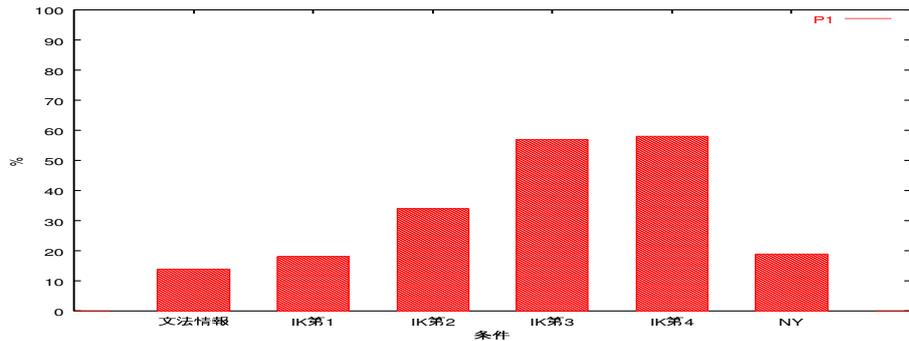


図 6: P1 の結果 (実験 2)

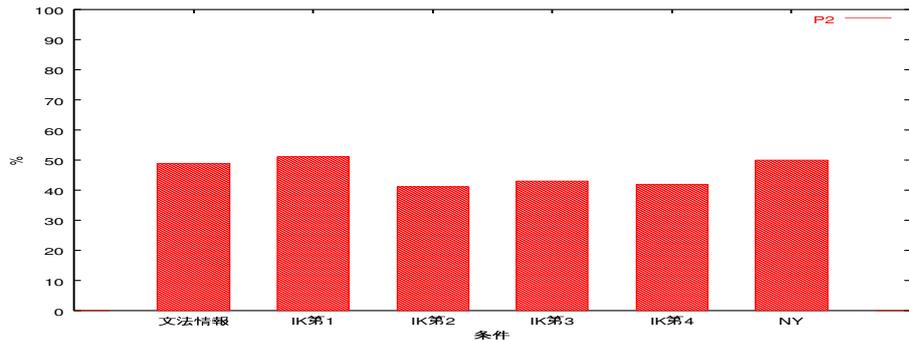


図 7: P2 の結果 (実験 2)

5 考察

本研究では、実験1および実験2を行ったが、文型パターンパーサおよび文型パターンファイルは実験2の方が最新版である(2006年2月時点)。よって、考察は実験2を元に行う。

5.1 意味属性の条件による実験結果の比較

実験より、各条件の結果を比較すると、IKの制約条件が厳しくなるほど適合率が下がり正解率が高くなる傾向が分かる。IK第2分類から第4分類は、詳細な意味の指定ができるように作られているが、詳細な意味を使用するほど適合率が大幅に下がる。

また、 $R1 \times P1$ の値はIK第3分類が最も高い。しかし、IK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。つまり、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第1分類を使用するのが最も適していると考えられる。

また、文法情報とIK第1分類の結果を比較すると、 $R1$ は文法情報の方が高い値を示しているが、 $P1$ および $P2$ についてはIK第1分類の方が高い値を示している。よって、IKは多少ではあるが正解率を上げる効果があったと言える。

5.2 IKとNYの結果比較

実験結果より、IK第1分類とNYを比較すると、適合率および正解率は近い値を示している。よって、より少ない数で意味を分類しているIK第1分類が文型パターンの絞りこみに適していると言える。

また、用言意味属性を用いた単語レベル文型パターンの絞りこみにおける適合率および正解率は、IK第1分類とNYの適合率および正解率の値が近い事より、本研究の結果から、単語レベル文型パターンにおける適合文型パターンの絞り込みにおいて、 $R1$ は44%、 $P1$ は19%、 $P2$ は51%程度が限界であると考えている。

5.3 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>

(日本語パターン)

/y \$1^{/tcfk N1 は} /tcfk N2 を /cf 頼って \$1 /ycf 上京した。

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. 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ので, 用言意味属性での絞り込み効果は確認できない. しかし, 字面での表記は用言の意味的制約としては最も厳しい制約であると考えられる.

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 パターンであり, 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, 字面での意味制約が最も効果的であると言える.

5.4 名詞意味属性の利用

本研究では、用言意味属性を用いて文型パターンの絞り込みを行ったが、池原意味体系は用言だけでなく名詞も意味を 371 に分類している [6]。名詞意味属性の例を以下に示す。

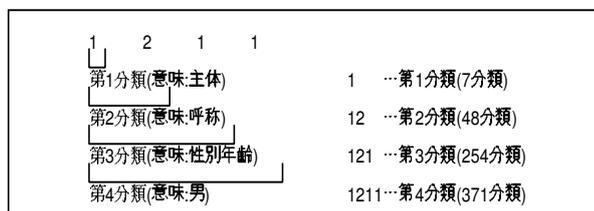


図 8: 名詞意味属性の例

今後は用言意味属性を用いた文型パターンの絞り込みだけでなく、名詞意味属性を用いた場合、動詞+名詞の意味属性を用いた場合など、条件を変える事による適合率および正解率を調査し、最適な意味属性を用いた文型パターンの絞り込み条件を明確にする必要がある。

5.5 句レベルおよび節レベルパターンの利用

文型パターンには単語レベルの他に句レベル, 節レベルがある. 句レベル及び節レベルの文型パターンは単語レベルの文型パターンに比べ, 句や節を変数化している. 例を以下に示す.

- <単語レベルパターン>
(日本語パターン)
 $/y \$1 /tcfk N1 \wedge /cf (\text{行く} | \text{いく}) ! \text{途中 } \$1^{\wedge} \{ /tck N2 \text{に} \} /cf V4.kako.$
- <句レベルパターン>
(日本語パターン)
 $/y \$1 /tcfk N1 \wedge /cf (\text{行く} | \text{いく}) ! \text{途中 } \$1^{\wedge} \{ /tck N2 \text{は} \} VP3.kako.$
- <節レベルパターン>
(日本語パターン)
 $/ytcfk N1 \wedge /cf (\text{行く} | \text{いく}) ! \text{途中 } /tck CL2.kako.$

5.3 節より, 字面での意味制約が最も効果的であると述べた. つまり, パターンの意味的制約が強いほど品質の高い翻訳ができる可能性が高いと考えられる. 文型パターン辞書には単語レベル, 句レベルおよび節レベルの文型パターンが収録されているが, 品詞をより細かく使用している点を考慮すると, 単語レベル, 句レベル, 節レベルの順で意味的制約が強い. よって, 単語レベルの文型パターンを使用した場合, 入力文に適した翻訳ができる可能性が最も高い. しかし, 単語レベルの文型パターンのみを使用した場合, 適合率および正解率を大幅に上げる事は難しい.

この問題を解決するために, 以下の方法が考えられる.

1. 単語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絞り込みを行う.
2. もし文型パターンが絞り込みで残らなかった, あるいは, 残った文型パターンを使用した翻訳結果が悪い場合. 句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う.
3. 句および節レベルで絞り込みで残ったパターンを使用して英文を作成する.

今後, 全てのレベルの文型パターンを使用する事で適合率および正解率は向上すると考えられる.

5.6 意味属性の付与誤り

本研究で使用した文型パターンの動詞部分に付与された意味属性に意味属性付与誤りが存在した。以下に例を示す。例中の原文とは、パターンを作成する際に使用された日本語文である。

- <意味属性の付与誤り例 1>

(日本語原文)

彼女は私を振って彼に走った。

(日本語パターン)

$/y\ \$1^{\wedge}\{/tcfk\ N1\ \text{は}\ } /tcfk\ N2\ \text{を}\ \$1\ /cf\ V3(\text{回転振動})(\text{て}\ |\ \text{で})\ \$1\ /ytck\ N4\ \text{に}\ /cf\ \text{走った。}$

意味属性の付与誤り例 1 において、V3 には意味属性「回転振動」が付与されているが、原文より V3 にはさらに意味属性「除去廃棄」を付与しなければならない。

- <意味属性の付与誤り例 2>

(日本語原文)

私は彼を頼って上京した。

(日本語パターン)

$/y\ \$1^{\wedge}\{/tcfk\ N1\ \text{は}\ } /tcfk\ \text{彼を}\ \$1\ /cf\ V2(\text{依頼要求})(\text{て}\ |\ \text{で})\ \$1\ /ycf\ (V3(\text{上方向移動}).kako|\ ND3\ \text{をした})。}$

意味属性の付与誤り例 2 において、V3 には意味属性「上方向移動」が付与されているが、原文より V3 にはさらに意味属性「行く出発」が付与されなければならない。

一方、抽出された全パターンからランダムに 25 パターン選択し、意味属性の付与誤りを調べた所、2 パターン (上記の 2 パターン) に付与誤りがあった。今後は、文型パターンの動詞部分に付与されている意味属性をチェックし、意味属性付与誤りを校正する必要がある。

5.7 各絞り込み条件で消えた例

実験(絞り込み条件:文法情報)で評価 A および B と評価された文型パターンが, 絞り込み条件によって消える問題がある. 例を以下に示す.

< IK 第 1 分類の絞り込みで消えた例 >

- <IK 第 1 分類の絞り込みで消えた例 1>

(入力文)

親に隠れ (IY:5320,IY:5950) て煙草を吸っ (IY:2100,IY:2220,IY:2233) た。

(日本語原文)

ポーチに座って風景を楽しんだ。

(英語原文)

We sat on the porch and enjoyed the scenery.

(日本語パターン)

WJAQ218660-00: /y </tk N1 は> /tcfk N2 に /cf V3(IY:2140,IY:2310,IY:8411)(て | で) /ytck N4 を /cf V5(IY:0200,IY:0211).kako.

(英語パターン)

WEAQ218660-00: <We| N1> V3.past on N2 and V5.past N4.

(出力英文 (評価))

I hid from the parents and somoked a cigarette. (B)

IK 第 1 分類の絞り込みで消えた例 1 における, 入力文の動詞「隠れる」用言意味属性は「現象事象の表現」であり, 「吸う」の用言意味属性は「日常生活の行為」である. また, 日本語パターンにおける V3 の用言意味属性は「日常生活の行為, 対物的行為」であり, V5 の用言意味属性は「知覚と情緒の表現」である. 入力文と日本語パターンの対応する動詞の用言意味属性を比較すると, 一致している用言意味属性は存在しない.

- <IK 第 1 分類の絞り込みで消えた例 2>

(入力文)

彼は 10 歳になる (IY:5310,IY:6000) までそこに住ん (IY:2231) でいた。

(日本語原文)

私の祖父は 70 歳になるまで農業に従事していた。

(英語原文)

My grandfather pursued the industry of farming until the age of seventy.

(日本語パターン)

WJBB058394-00: /y \$1^{/tcfk #1[N2 の] /k N3 は} /tcfk NUM4 に /cf なる (ま | 迄) \$1 /yc N5 に \$1 /cf (V6(IY:4630).teiru.kako| ND6 をしていた)。

(英語パターン)

WEBB058394-00:#1[N2^poss] N3 V(V6| ND6).past N5 until NUM4. (出力英

文 (評価))

I lived there until ten. (A)

同様に, IK 第 1 分類の絞り込みで消えた例 2 における, 入力文の動詞「なる」用言意味属性は「現象事象の表現, 変化の現象事象の表現」であり, 「住む」の用言意味属性は「日常生活の行為」である. また, 日本語パターンにおける V6 の用言意味属性は「社会的活動の行為」である. 入力文と日本語パターンの対応する動詞の用言意味属性を比較すると, 一致している用言意味属性は存在しない.

- <IK 第 1 分類の絞り込みで消えた例 3>

(入力文)

母は赤ん坊をあやし (IY:0351) て笑わせ (IY:0342) た。

(日本語原文)

泥棒は靴を脱ぎ捨てて逃げた。

(英語原文)

The thief flung off his shoes and fled.

(日本語パターン)

WJAF020728-00: /y \$1^{/tcfk N1 は} /tcfk N2 を \$1 /cf V3(Y:2235)(て | で) \$1 /ycf V4(IY:3420,IY:7520).kako。

(英語パターン)

WEAF020728-00: N1 V3.past N1^pron^poss N2 and V4.past.

(出力英文 (評価))

Mother played her baby and laugh. (B)

同様に, IK 第 1 分類の絞り込みで消えた例 3 における, 入力文の動詞「あやす」と動詞「笑わせる」用言意味属性は「知覚と情緒の表現」である. また, 日本語パターンにおける V3 の用言意味属性は「日常生活の行為」であり, V4 の用言意味属性は「地域社会生活の行為, 移動行為の表現」である. 入力文と日本語パターンの対応する動詞の用言意味属性を比較すると, 一致している用言意味属性は存在しない.

< IK 第 2 分類の絞り込みで消えた例 >

- <IK 第 2 分類の絞り込みで消えた例 1>

(入力文)

私は彼を頼っ (IY:2440) て上京し (IY:7210) た。

(日本語原文)

彼はまっさらなスーツを着て来た。

(英語原文)

He came wearing a shiny new suit.

(日本語パターン)

WJBD017868-00: /ytcfk N1は#1[/cf AJV3^rentai] /f N4を /cf V5(IY:0120,IY:2235, IY:4320,IY:8312)(て | で) /yef V6(IY:5310,IY:7420).kako.

(英語パターン)

WEBD017868-00:N1 V6.past AJ(V5) #1[shiny AJ3] N4.

(出力英文 (評価))

I came to Tokyo count on he. (B)

IK 第2分類の絞り込みで消えた例1における、入力文の動詞「頼る」の用言意味属性は「対人的な行為」であり、動詞「上京する」用言意味属性は「上下の移動」である。また、日本語パターンにおけるV5の用言意味属性は「知覚感覚, 家庭生活の行為, 規則契約」であり、V6の用言意味属性は「発生消滅, 行き来の移動」である。入力文と日本語パターンの対応する動詞の用言意味属性を比較すると、一致している用言意味属性は存在しない。

- <IK 第2分類の絞り込みで消えた例2>

(入力文)

母は赤ん坊をあやし (IY:0351) て笑わせ (IY:0342) た。

(日本語原文)

ストローを差しこんで飲んだ。

(英語原文)

She inserted the straw and drank.

(日本語パターン)

WJAJQ240972-00: /y </tk N1は> /tcfk N2を /cf V3(IY:0130,IY:7120,IY:8311)(て | で) /yf V4(IY:0342,IY:2100,IY:2233,IY:2450).kako.

(英語パターン)

WEAQ240972-00: <She|N1> V3.past N2 and V4.past.

(出力英文 (評価))

Mother played the baby and laugh. (B)

同様に、IK 第2分類の絞り込みで消えた例2における、入力文の動詞「あやす」及び「笑わせる」の用言意味属性は「対人的感性」である。また、日本語パターンにおけるV3の用言意味属性は「知覚感覚, 出入りの移動, 対象依存の行為」であり、V4の用言意味属性は「対人的感性, 身体的動作, 対人的な行為」である。入力文と日本語パターンの対応する動詞の用言意味属性を比較すると、「笑わせる」とV4で意味属性が一致しているが、「あやす」とV3の意味属性が一致していない。

- <IK 第2分類の絞り込みで消えた例3>

(入力文)

彼は彼女を振り捨て (IY:8423) て他の女に走っ (IY:2130,IY:7000,IY:7520) た。

(日本語原文)

彼女は私を振って彼に走った。

(英語原文)

She jilted me for him.

(日本語パターン)

WJAM044111-00: /y \$1^{/tcfk N1は} /tcfk N2を \$1 /cf V3(IY:0252,IY:2120,IY:8120, IY:8210)(て|で) \$1 /ytck N4に /cf 走った。

(英語パターン)

WEAM044111-00:N1 V3.past N2^obj for N4^obj.

(出力英文(評価))

He discarded her for her. (B)

同様に、IK 第2分類の絞り込みで消えた例3における、入力文の動詞「振り捨てる」の用言意味属性は「創造破壊の行為」であり、動詞「走る」用言意味属性は「身体的動作、移動行為の表現、遠近移動の事象」である。また、日本語パターンにおけるV3の用言意味属性は「個人的感情、身体的動作、単一的物体への行為」である。入力文と日本語パターンの対応する動詞の用言意味属性を比較すると、一致している用言意味属性は存在しない。

< IK 第3分類の絞り込みで消えた例 >

- <IK 第3分類の絞り込みで消えた例1>

(入力文)

先生は生徒の学力に合わせ (IY:1425,IY:4110,IY:5820,IY:5930,IY:5940,IY:8340)
た試験問題を作っ (IY:1510,IY:2210,IY:4710,IY:4720,IY:5310) た。

(日本語原文)

彼はテープに録音した歌を消した。

(英語原文)

He erased the songs recorded on the tape.

(日本語パターン)

WJBD026841-00: /y \$1^{/tcfk N1は} /tcfk N2に /cf (V3(IY:8300).kako^rentai|ND3
をした) /f N4を \$1 /cf V5(IY:5320,IY:5520,IY:8320).kako。

(英語パターン)

WEBD026841-00:N1 V5.past N4 V(V3|ND3).past on N2.

(出力英文(評価))

The teacher made a test apprpriated for the student. (B)

IK 第3分類の絞り込みで消えた例1における、入力文の動詞「合わせる」の用言意味属性は「研究調査、結成拡散、拡散集合、適合不適合、均衡過不足、液体」であり、動詞「作る」用言意味属性は「創作活動、生死生育、農林業漁業、製造業、発生」である。また、日本語パターンにおけるV3の用言意味属性は「対象依存の行為」であり、V5の用言意味

属性は「消滅, 光, 火」である. 入力文と日本語パターンの対応する動詞の用言意味属性を比較すると, 一致している用言意味属性は存在しない.

- <IK 第3分類の絞り込みで消えた例2>

(入力文)

彼にはその任務を果たせる (IY:0263,IY:1421,IY:5970) だけの能力がなかった。

(日本語原文)

それをするだけの勇気がなかった。

(英語原文)

I didn't have the bravery to do it.

(日本語パターン)

WJAQ029011-00: /y </tk N1は> /tcfk N2を /cf V3(IY:1450,IY:2100,IY:2244,IY:2300, IY:2320,IY:4510,IY:4770,IY:5910,IY:5913,IY:7720,IY:8312,IY:8360,IY:8421, IY:8423)^rentai だけの! N4が /cf (無かつ|なかつ)た。

(英語パターン)

WEAQ029011-00: <I|N1> did not have N4 to V3 N2.

(出力英文 (評価))

He did not have the ability to carry out. (B)

同様に, IY 第3分類の絞り込みで消えた例2における, 入力文の動詞「果たす」の用言意味属性は「願望努力忍耐, 研究調査, 順序と時間」である. また, 日本語パターンにおける V3 の用言意味属性は「まとめと発表, 身体的動作, 生活一般, 個人的な行為, 所有保存, 通知報道, 編集出版興業, 相互関係, 受領的移動, 物体表面, 紙上物体, 分離破壊」である. 入力文と日本語パターンの対応する動詞の用言意味属性を比較すると, 一致している用言意味属性は存在しない.

- <IK 第3分類の絞り込みで消えた例3>

(入力文)

私の弁護士が私に代わっ(IY:7730)て合意書を作成し(IY:1424,IY:4720,IY:5310,IY:5920)てくれた。

(日本語原文)

弁護士が私に代わって問題を解決してくれた。

(英語原文)

My lawyer straightened the matter out for me.

(日本語パターン)

WJAQ240693-00: /y #1{/tcfk N1が, /tcfk N2に} /cf 代わって /ytek N3を /cf V3(IY:1434)#[.tekureru].kako.

(英語パターン)

WEAQ240693-00:N1 V3 N3-v3obj for N2^obj.

(出力英文 (評価))

The lawyer drew up an agreement for me. (B)

同様に, IK 第3分類の絞り込みで消えた例3における, 入力文の動詞「代わる」の用言意味属性は「交換的移動」であり, 動詞「作成する」用言意味属性は「研究調査, 製造業, 発生, 要不要成立不成立」である. また, 日本語パターンにおけるV3の用言意味属性は「分類評価」である. 入力文と日本語パターンの対応する動詞の用言意味属性を比較すると, 一致している用言意味属性は存在しない.

< IK 第4分類の絞り込みで消えた例 >

- <IK 第4分類の絞り込みで消えた例>

(入力文)

彼にはその任務を果たせる (IY:0263,IY:1421,IY:5970) だけの力がなかった。

(日本語原文)

彼には自分の意見を押し通すだけの力がなかった。

(英語原文)

He had little power to enforce his ideas.

(日本語パターン)

WJ AQ123080-00: /ytcfk N1には#1[/cf 自分の] /k N3を /cf V4(IY:0265,IY:2430)^rentai だけの! N5が /cf (無かつ| なかつ) た。

(英語パターン)

WE AQ123080-00:N1 had little N5 to V4 #1[N1^poss] N3.

(出力英文 (評価))

He had little ability to carry out the task. (B)

IK 第4分類の絞り込みで消えた例における, 入力文の動詞「果たす」の用言意味属性は「努力忍耐, 実施実行, 順序と時間」である. また, 日本語パターンにおけるV4の用言意味属性は「奮起執着, 約束交渉」である. 入力文と日本語パターンの対応する動詞の用言意味属性を比較すると, 一致している用言意味属性は存在しない.

最も制約が緩いIK 第1分類を用いた文型パターンの絞り込みにおいて, AおよびBと評価された文型パターンが60パターン削除された. 削除された文型パターンと入力文を調べたところ, 入力文と文型パターンの対応する動詞の意味が近いものは存在しなかった. よって, 用言意味属性の意味分類を再構成しても, 入力文に適合する文型パターンが増える可能性は低いと考えられる.

6 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することを目的とした。実験の結果、実験1では $R1$ 、 $P1$ および $P2$ において最も高い値はそれぞれ、 $R1$ は 33%(文法情報)、 $P1$ は 57%(IK 第4分類)、 $P2$ は 55%(文法情報)であった。また、実験2では $R1$ 、 $P1$ および $P2$ において最も高い値はそれぞれ、 $R1$ は 0.53(文法情報)、 $P1$ は 0.58(IK 第4分類)、 $P2$ は 0.51(IK 第1分類)であった。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。しかし、適合率および正解率は高いとは言えない。

また、IK 第1分類を用いた文型パターンの絞り込みにおいて、A および B と評価された文型パターンが 60 パターン削除された。IK は、NY に比べ、意味の偏りが少なくなっている。文型パターンに付与された IK には多少の用言意味属性付与誤りがあるが、IK 用言意味属性第1分類には意味の偏りが無い事から、文型パターンの意味に偏りがあるとは考えづらい。よって、IK の意味分類を再考し構築し直しても、IK 第1分類を用いた絞り込みで削除された入力文に適合する文型パターンを用言意味属性で拾う事ができる可能性は低いと考える。しかし、用言意味属性付与誤りを校正した文型パターンを用いて実験を行う必要はある。

今後はさらに適合率および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。また、名詞意味属性を用いたパターンの絞り込み、名詞+動詞の意味属性を用いたパターンの絞り込みを行い、適合率を低下させず、正解率を上げる為に最適な意味属性の条件を明確にする予定である。

謝辞

本研究は、科学技術振興事業団「JST」の戦略的基礎研究推進事業「CREST」における研究領域「高度メディア社会の生活情報技術」の研究課題「セマンティックタイポロジーによる言語の等価変換と生成技術」の支援により行いました。

本論文作成に際して、多大なる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:非線形な言語表現と文型パターンによる意味の記述, 情報処理学会, 自然言語処理研究会,2004-NL-159,pp.139-146,2004-1.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 長尾ほか:「自然言語処理」,ISBN4-00-010355-5, 岩波書店,1996.
- [6] 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.
- [7] 池原ほか:「日本語語彙大系」,ISBN4-00-130101-6, 岩波書店,1997.