

平成14年度 修士論文

意味属性の共起による  
「AのB」型名詞句の翻訳規則

Translation Rules for Japanese Noun Phrase “ A no B ”  
based on collocating pattern of semantic attribute

指導教官 池原 悟 教授  
村上 仁一 助教授  
徳久 雅人 助手

鳥取大学大学院 工学研究科 知能情報工学専攻  
M01T2031C 守谷 有司

## 内容梗概

日本語において、よく用いられる「AのB」型名詞句の翻訳規則を意味属性 [1] の共起パターンを用いて記述し、実験を行うことによってその精度について検討をする。具体的には、まず英語辞典等から対訳を収集したデータベース [7] から「AのB」型名詞句の標本を約3万件抜き出した。抜き出した標本は、英語表現による着目をしたところ49種類に分類できた。翻訳規則は、出現頻度が10件以上の英語表現28種類を対象とした。次に、意味属性規則の学習データとして、標本5000件に対し意味属性の付与を行った。意味属性の付与は、文意に合致する意味属性のみを付与するため、標本を形態素解析プログラムALT-JAWSに掛けた後、人手で絞り込みを行った。また、相補的規則として字面と動詞の規則も作成した。得られた翻訳規則をオープンデータ1000件を用い計算機による実証をしたところ、意味属性のみの規則による翻訳精度は約55%であり、字面・動詞規則を併用した翻訳規則の精度は約65%であった。しかし、形態素解析による失敗を除くと正解率はそれぞれ約59%、約70%であった。他に、文脈によって正解となる英語表現が19%、20%と比較的多く、明確な誤りとなった英語表現が21%、17%となった。

# 目次

内容梗概	i
<b>第 1 章 序論</b>	<b>1</b>
1.1 はじめに	1
1.2 英語表現について	1
1.3 従来研究及び本研究の目的	3
<b>第 2 章 翻訳規則の作成</b>	<b>4</b>
2.1 一般名詞意味属性	4
2.2 意味属性規則の作成	5
2.3 標本の収集	5
2.3.1 「A の B」型名詞句の抜粋	7
2.3.2 抜き出した対訳の詳細	9
2.4 英語表現の分類	10
2.5 学習データの作成	11
2.5.1 意味属性の付与	12
2.6 規則の生成	13
2.6.1 汎化の原理	13
2.6.2 汎化の効果	13
2.6.3 汎化のアルゴリズム	14
2.6.4 意味属性規則の適用について	18
2.7 相補的規則の作成	19
2.7.1 字面規則の作成	19
2.7.2 動詞規則の作成	20
<b>第 3 章 実験</b>	<b>23</b>
3.1 実験 1	24
3.1.1 考察	25
3.2 実験 2	30
3.2.1 考察	30
3.2.2 意味属性規則の分布について	32
3.3 実験 3	34
<b>第 4 章 検証</b>	<b>35</b>
4.1 学習データの検証	35
4.2 汎化・適用順序の検証	36
4.2.1 汎化回数の上限	36

4.2.2 汎化順序の変更 . . . . .	37
<b>第 5 章 結論</b>	<b>38</b>
5.1 おわりに . . . . .	38
5.2 今後の課題 . . . . .	38
<b>参考文献</b>	<b>40</b>

## 図版目次

2.1	意味属性体系 . . . . .	4
2.2	対訳抜粋方法 . . . . .	8
2.3	収集した標本の英語表現分布 . . . . .	10
2.4	両者の意味属性 . . . . .	13
2.5	意味属性体系のグループ化 . . . . .	14
2.6	汎化1次元 . . . . .	14
2.7	汎化2次元 . . . . .	15
2.8	上位意味属性の探索 . . . . .	18
2.9	字面規則の書式 . . . . .	19
2.10	動詞規則の書式 . . . . .	20
3.1	意味属性規則の分布図 . . . . .	33
4.1	汎化の上限 . . . . .	36

# 表目次

1.1	英語表現 . . . . .	2
1.2	従来の研究において翻訳対象としていた英語表現 . . . . .	3
2.1	対訳抜粋に失敗した句の原因 . . . . .	9
2.2	意味属性規則の作成数 . . . . .	16
2.3	意味属性規則 . . . . .	17
2.4	字面規則 . . . . .	21
2.5	動詞規則 . . . . .	22
3.1	評価基準 . . . . .	25
3.2	オープンテストデータ 1000 文の内訳 . . . . .	25
3.3	実験 1 の結果 . . . . .	26
3.4	実験 2 の結果 . . . . .	31
3.5	実験 3 の結果 . . . . .	34
4.1	学習データ検証結果 . . . . .	35
4.2	汎化回数の変化に伴う精度の遷移 . . . . .	36
4.3	汎化順序の変更に伴う正解率の変化 . . . . .	37

# 第 1 章 序論

## 1.1 はじめに

日本語において、名詞と名詞を「の」で結んだ「A の B」という型の名詞句はよく用いられる表現であり、多彩な英語表現を持つ。従って、この「A の B」型名詞句の翻訳規則を作成することで、日英機械翻訳の翻訳精度向上につながると考えられる。

本研究室でも、これまで名詞間の意味関係に着目した翻訳規則について検討してきた(文献 [3] 参照。以後従来研究とする)。しかし従来の研究では、翻訳対象とする英語表現を 8 種類に絞った検討しかなされておらず、また計算機による実証もなされていない。

そこで本研究では、「A の B」型名詞句の翻訳規則を意味属性 [1] の共起パターンを用いて計算機処理可能レベルで作成し、計算機による実証を行う。また、多くの標本を収集することにより、訳出する英語表現を広げる。

## 1.2 英語表現について

本研究で扱う英語表現とは、日本語の「A の B」型名詞句に対する英訳において、A と B の間にある助詞を基に、英語の型を一般化したものである。以下に例文を示す。

例文：彼は、一人で滝のしぶきを浴びながら立っていた。

対訳：He was standing alone in the spray of the waterfall.

例文では「滝のしぶき」に対応する英訳部分は“ the spray of the waterfall ”である。A 名詞の「滝」に対応する“ the waterfall ”と、B 名詞の「しぶき」に対応する“ the spray ”の間の助詞“ of ”を基に、この英訳の英語の型を一般化すると「B of A」となり、これを英語表現とする。表 1.1 に英語表現の例を示す。

表 1.1: 英語表現

日本語原文	英訳	英語表現
店の事務員	store as a clerk	A as B
衣服の卸	clothes at wholesale	A at B
日曜の朝	Sunday morning	A + B
生徒の顔	the students by sight	A by B
従業員の給料	the employees' salary	A' B
市長の怠慢	the mayor for his negligence	A for B
野菜の腐敗	vegetables from decay	A from B
敵の背後	the enemy in the rear	A in B
カメラの焦点	camera into focus	B into A
部屋のすみ	the corner of the room	A of B
子供の頬	child on the cheek	A on B
川の水源	river to its source	A to B
孤児の保護	orphan under my protection	B under A
風呂の水	bath-tub with water	A with B
どちらの方向	what direction	疑問詞+B
他の区域	other area	形容詞 + B
去年の今頃	this time last year	B + A
遊覧船の乗客	passengers aboard the pleasure boat	B aboard A
人々のうわさ	gossiping about her neighbors	B about A
極地の大気	atmosphere above the polar	B above A
湖水の見晴らし	outlook across the lake	B across A
雪の日	day after snow	B after A
火の用心	precautions against fire	B against A
毎日の雨	raining almost every day	B almost A
海岸の松	pinos along the shore	B along A
学生の尊敬	respect among his student	B among A
国中の人々	people around the country	B around A
ブレーキの役	act as a brake	B as A
宴のスピーチ	the speeches at the wedding reception	B at A
汽車の旅	traveling by train	B by A
町の境	boundaries between town	B between A
中世のヨーロッパ	Europe during the Middle Ages	B during A
学芸会のプログラム	a program for our cultural festival	B for A
井戸の水	water from a well	B from A
雪国の風景	scenery in the snow country	B in A
天候の変化	change of weather	B of A
ラジオの音量	the volume on a radio	B on A
島の管轄	jurisdiction over the island	B over A
牧師の職業	profession such as pastor	B such as A
トンネルの半ば	Halfway through the tunnel	B through A
世界の学界	academic circles throughout the world	B throughout A
疑惑の人	man under suspicion	B under A
ノイズの影響	susceptibility to noise	B to A
半袖のシャツ	an undershirt with half sleeves	B with A
組織の秩序	order within the organization	B within A
ストーブの回り	around the stove	前置詞 + A
ぼくのおばあさん	My grandmother	所有格 + B
プロセッサーのアタッチメント	Processor's Attachments	A's + B
映画のファン	moviefan	単語



### 1.3 従来研究及び本研究の目的

従来研究 [3] では、表 1.2 に示す英語表現について、A・B間の意味関係に着目して、A・B間の意味的關係を定義することにより「AのB」型名詞句の英語表現を決定できるかについて検討を行っていた。

具体的には、“BがAの所属関係にある”等の7種類の意味関係を定義し、その意味関係及び意味属性・文法属性を用いて以下のように72種類の翻訳規則を作成した。

- BがAの場を表す場合、英語表現は「B of A」
- BがAの所属である場合、英語表現は「B of A」
- Aの意味属性が【学問名】、Bの意味属性が【研究】【勉強】である場合、英語表現は「B of A」

この規則を1000件のオープンデータを用い、机上による評価を行ったところ、正解率は約63%ととなり、作成した8種類の英語表現に対して翻訳規則を適用したところ約76%の結果が得られた。

しかし、“BがAの一部”等の意味的關係を機械規則として記述することは困難である。そのため、完全に機械処理可能な規則として記述できず、机上による評価しか行われていなかった。また、評価によると17.2%の標本が作成した翻訳規則の英語表現の対象外となった。

そこで本稿では、意味属性の共起を用いることにより完全に計算機処理可能なレベルで翻訳規則を作成し、対象とする英語表現を広げることにより、規則のカバー率を上げることがを目的とする。

以上1章では、従来研究における翻訳規則の問題点、及び本稿の研究目的について述べた。2章では、翻訳規則の作成方法について述べ、3、4章で作成した翻訳規則を実験により評価し、その考察及び検討を行う。

表 1.2: 従来研究において翻訳対象としていた英語表現

英語表現	
A's + B	B in A
B of A	単語
A + B	B + A
形容詞+B	B for A

## 第 2 章 翻訳規則の作成

従来研究 [3] より，A・B間の意味的な関係を用いることにより「AのB」型名詞句の英語表現の翻訳ができることが予想できる．そこで意味属性の共起パターンを用いて翻訳規則を作成する．

### 2.1 一般名詞意味属性

一般名詞意味属性とは，一般名詞について概念化された対象と単語との対応関係を，対象の見方，捉え方に着目して整理分類されたものである．意味属性体系は図 2.1 のように最大 12 段 (属性数，2,710) の木構造になっており，上位のノードの性質を下位のノードが継承するという性質を持っている．

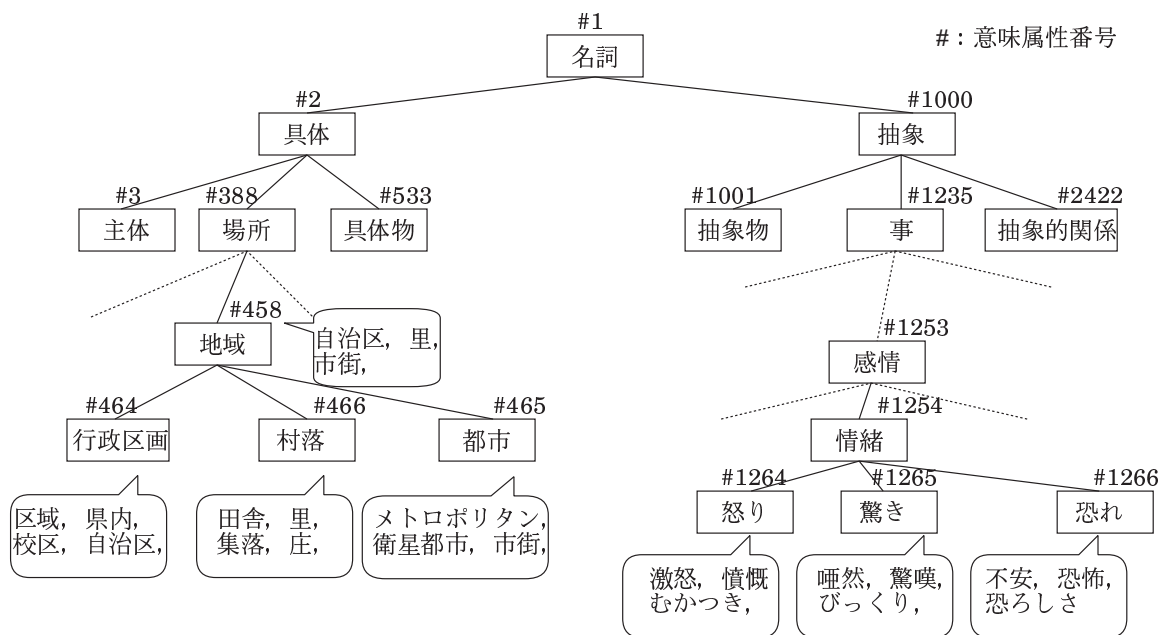


図 2.1: 意味属性体系

## 2.2 意味属性規則の作成

意味属性規則の作成手順は次の通りである。

1. 標本の収集
2. 英語表現の分類
3. 学習データの作成
4. 規則の生成

以下に具体的に説明していく。

## 2.3 標本の収集

翻訳規則を作成するためには、多くの標本が必要不可欠である。現在、「AのB」型名詞句の対訳コーパスは存在しない。そこで、英語辞書などから作成した対訳データベース(文献[7]、以下対訳データベースとする。)30万件のデータからまず「AのB」名詞句に該当する箇所を大雑把に計算機で抜きだし、後に手作業で整備する。翻訳対象とする「AのB」名詞句は、係り受け関係があるものとし、複合名詞などは、対象としない。以下に対象とする名詞句についての注意点について述べる。

- 取り扱う「AのB」型名詞句
  1. 名詞 A と B に係り受け関係がある名詞句
 

例文 1：その野球の試合はテレビで放映された。  
The baseball game was telecast(ed).

例文 2：暴動の危険は遠のいている。  
The risk of riots is receding.
  2. 用言性一般名詞名詞
 

例文 3：展覧会の案内を出す。(用言性名詞)  
To send out invitations to an exhibition.

例文 4：山田さんがこのアパートの管理をしている。(～ing型)  
Mr. Yamada is supervising this apartment building.
- 除外する名詞句
  1. 意訳(対応する英訳部分が名詞句ではない)
 

例文 5：車がいっぱいで道路の交通がまひした。  
Cars clogged the road.

例文 6：彼女の服装は地味だ。  
She was dressed quietly.

例文 7：彼女はいくらか医学の心得がある。  
She knows something of medicine.

例文 8：食事の支度が出来ましたから、席に着いて下さい。  
When the meal is ready, please come to your seat.
  2. 英訳部分が省略されている場合
 

例文 9：彼は、骨の随まで冷え切っていた。  
He was chilled to the bone.

例文 10：ポストの所を右に曲がって5軒目が私の家です。

Turn right at the mailbox and the fifth door down is mine.

例文 11：たばこの火で畳が焦げてしまった。

The tatami was burned by the cigarette.

3. 比喩

例文 12：今年は梅雨の入りが遅い。

The rainy season is late coming this year.

例文 13：明日は明日の風が吹く。

Tomorrow brings its own fortune.

4. 複合名詞

例文 14：競馬専門のチャンネル

a cable channel dedicated to horse racing

5. 接頭，接尾辞を含む句

例文 15：初対面の挨拶。

Greetings on a first meeting.

例文 16：切符を買うために 100人以上の人が長い行列を作っていました。

More than one hundred people had formed a long line to buy tickets.

6. 形式名詞

例文 17：寒い冬のために、農家では、野菜を囲っておきます。

The farmers are preserving their vegetables because of the cold winter.

例文 18：口癖の様に言う。

To be in the habit of saying.

例文 19：私は洋服よりも着物の方が合う。

Kimono suits me better than western clothes.

7. 数詞

例文 20：この講習は、8週間の予定です。

This course is planned for eight weeks.

例文 21：私の組は、火曜日の1時間目が「日本歴史」です。

My class has "Japanese History" in the first period on Tuesday.

8. その他

例文 22：生命や財産の安全を守る。(N や N の N)

To protect the safety of lives and property.

例文 23：駅前のビルの二階が僕の事務所だ。(N の N の N の N)

My office is on the second floor of the building in front of the station.

例文 24：挨拶の仕方は、家庭で母に教えられました。(仕方：how to)

I was taught how to make proper greetings by my mother at home.

### 2.3.1 「A の B」型名詞句の抜粋

抜粋方法を図 2.2 に示し、手順を以下に示す。

1. 対訳データベースを形態素解析プログラム ALT-JAWS にかける。
2. 日本語「A の B」名詞句を検索し各名詞 A, B を読み込む。
3. 辞書テキストから A, B の英単語を検索する。検索の際、「A の B」のように単語として登録されている名詞「A の」のように形容詞型で登録されている名詞「A」のように名詞として登録されている名詞がある。辞書テキストにこれらの型が複数存在する場合は、プライオリティを高い順に「A の B」「A の」「A」という順に定め、該当する英訳を検索する。
4. 検索してきた英単語を基に対訳から A・B を検索し、A・B の英単語が対訳に共に存在すれば、A・B(B・A)間の箇所を抜き出す。

この結果、日本語においては約 84,000 件の名詞句 (複合語を含む) を収集した。また、対訳も抜け出せた名詞句は、約 30,000 件の名詞句であった。しかし上記の方法では誤った箇所の対訳を抜き出す場合もある。そこで抜き出した 30,000 件の名詞句について人による整備を行ったところ、約 3,500 件の誤った対訳を抜き出した。従って、収集数は、約 27,000 万件となった。

入力文

金の問題は当事者どうしで解決すべきだ。  
The question of money should be settled by the parties concerned.

⇩ (1).形態素解析

金の問題は当事者どうしで解決すべきだ  
1. 金(1100,[715,710])(1100,[1190,934])/の(7410)  
2. 問題(1100,[1033,1511])/は(7530)  
3. 当事者(1100,[341])/どうし(6280,[2602,2660],間)/で(7410)  
4. 解決す(2436,[1447],解決する)/べき(7297)/だ(7256)  
-----  
The question of money should be settled by the parties concerned  
-----

(2)

A: 金, B: 問題

(3)検索

辞書テキスト

■ 金 : Benjamin ● folding money  
[cabbage, green, lettuce, stuff] ● auri ● auro…  
● ducat ● dust ● golden calf…  
● main line ● mean green ● mon ● money ● mulla  
● nicker ● oday ● oofus…  
● wampum [【語源】 wampumpeagの短縮形]  
● whip-out ● yellow metal…  
■ 金 {人名-2} : Kim [朝鮮、韓国、北朝鮮。]  
■ 金・外貨準備 : reserve of gold and foreign currency  
⋮  
■ 問題 : affair ● case ● defect ● difficulty ● direction  
● equation ● issue ● problem ● quare ● question  
● trouble ● woes ● wrinkle  
■ 問題 ; ~についての : question about ● question as to  
■ 問題 ; ~に関する : concern about ● issues related to  
■ 問題 ; ~の : matter of  
■ 問題 ; ~をめぐる : problem over

(4)参照

The question of money should be settled by the parties concerned.

⇩  
金の問題 : The question of money

図 2.2: 対訳抜粋方法

### 2.3.2 抜き出した対訳の詳細

2.3.1 節における抜粋方法では、抜粋できない名詞句がある。そこで抜粋に失敗したデータからランダムに 1,000 件抜きだし、1,000 件のデータの性質を、失敗する原因を基に調査する。

調査の結果、失敗する原因は表 2.1 のように分類できた。

表 2.1: 対訳抜粋に失敗した句の原因

失敗原因	件数 (件)
意識 (対訳が名詞句になっていない)	548 件
辞書登録されていない単語のため	113 件
複合語	110 件
形式名詞	47 件
数詞	29 件
英訳省略	27 件
形態素解析失敗	21 件
比喻	18 件
その他	87 件

この結果より、“辞書登録されていない単語のため”失敗した場合等以下に示すものがまだ抜粋可能であることが分かる。

- 辞書登録されていない単語のため失敗  
 例文 1: アンダーラインの箇所を訳しなさい。  
 Translate the underlined parts.  
 例文 2: 彼は医師の治療を受けている。  
 He is under medical treatment.  
 例文 3: 都会では季節の移り変わりを楽しむことができない。  
 People in big cities cannot enjoy the changes of the seasons.
- 複数型による失敗 (その他)  
 例文 4: 工場の煙で空気が汚れている。  
 The air is polluted by smoke from the factories.

そこで、標本の偏りを減少させるため、2.3.1 節で失敗した文 10,000 件に対し、人手で対訳を抽出する。結果 2,000 件強の名詞句を抽出した。その結果、最終的な標本数は約 3 万件となった。

## 2.4 英語表現の分類

翻訳対象とする英語表現を明確にし、また英語表現に関する規則を作成するために、標本を英語表現ごとに分類する。結果、標本は 49 種類の英語表現に分類できることが分かった。ところで、英語表現の中には、全体から見て出現頻度が極めて低い英語表現も多数存在し、学習データとして十分な標本が確保できない英語表現もある。ゆえに、本研究では出現頻度が 10 件以上である 28 種類の英語表現を研究対象とする。これは今回収集した標本全体の 99 % に相当する (図 2.3)。

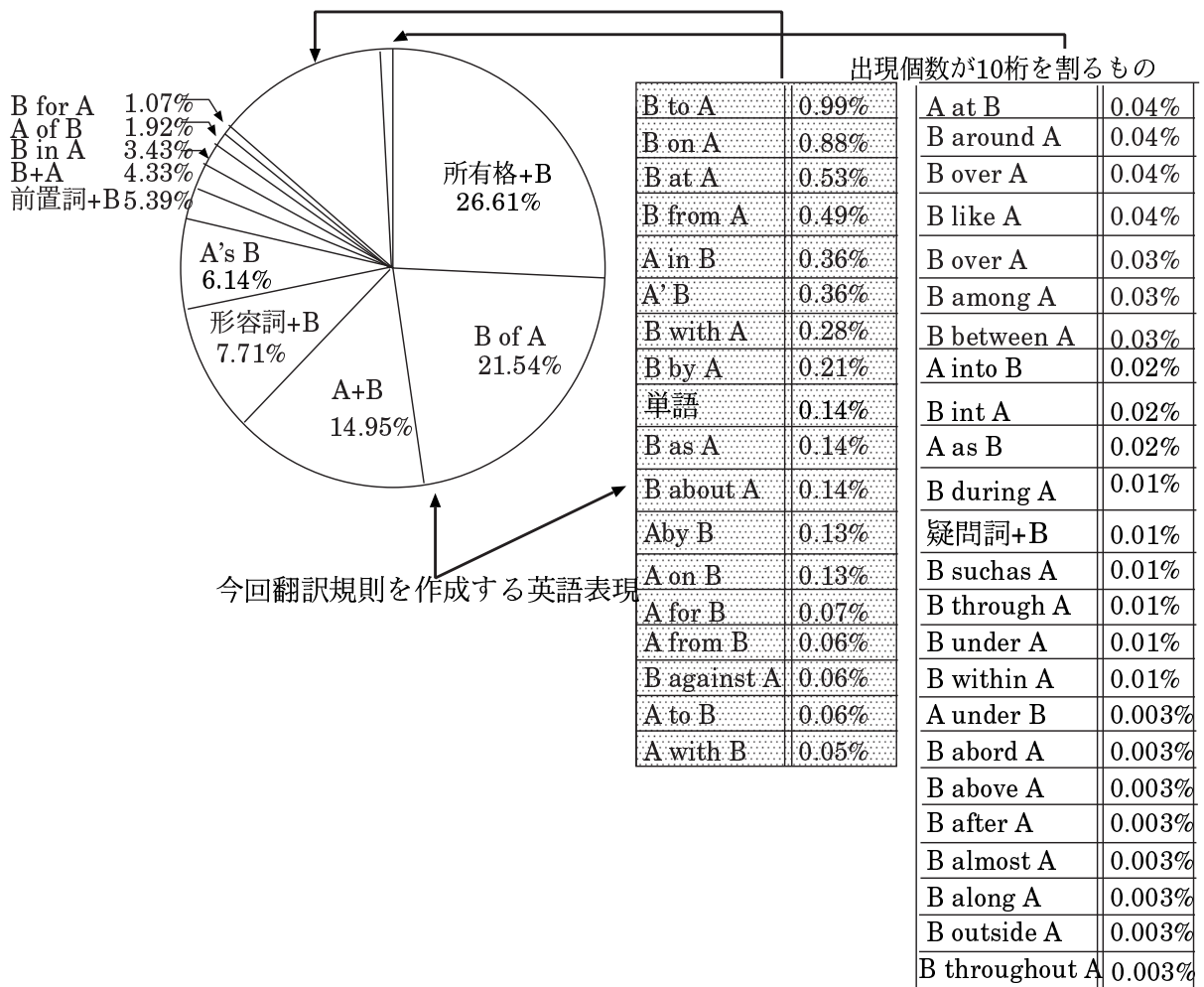


図 2.3: 収集した標本の英語表現分布



## 2.5 学習データの作成

標本の一部を学習データとして使用することにより意味属性規則を作成する。ここで、意味属性規則を作成するためには、学習データに意味属性が付与されている必要がある。付与は、形態素解析プログラム ALT-JAWS により行えるが、通常名詞は多義を持っているため、一名詞に付与される意味属性は一つではない。「A の B」の意味属性の組を学習データとする場合、「A の B」の意味属性には文意にあう意味属性が付与されていなければ意味を成さない。

例として、「インクのしみ」という名詞句を考える。

例：インクのしみ

原文：消しゴムでインクのしみを消そうとしたが消せなかった。

対訳：I tried to erase an ink spot,but in vain(A+B).

「インクのしみ」を形態素解析にかけると次のようになる。

形態素解析後の「インクのしみ」の箇所：

インク (#917【文具】) の

しみ (#2329【汚染】 #628【ほくろ・いぼ】 #2584【模様】)

形態素解析にかけたデータを学習データとして使用すると

A が #917【文具】 B が #2329【汚染】                    ならば英語表現“ A+B ”型

A が #917【文具】 B が #628【ほくろ・いぼ】        ならば英語表現“ A+B ”型

A が #917【文具】 B が #2584【模様】                ならば英語表現“ A+B ”型

という 3 つの規則が生成される。しかし、実際文で用いられている意味は“【文具】の【汚染】”であり、“【文具】の【ほくろ・いぼ】”、“【文具】の【模様】”ではない。従って、“【文具】の【ほくろ・いぼ】”や“【文具】の【模様】”も規則として使用すれば、意味属性を用いる意味が無くなる。そこで、学習データには文意に合致する意味属性の付与を行う。

### 2.5.1 意味属性の付与

現在のところ，文意に合う意味属性を一意に決定させる方法がないため，収集した標本 5000 件に対して人手で意味属性の付与を行う．具体的には，形態素解析プログラムにかけたデータを人手で一つずつ処理する．尚，一意に決定できない名詞も存在するため，これらの名詞については複数の意味属性を付与する．以下に意味属性付与の際の例を示す．尚，例中の # は意味属性番号【】は意味属性名，b は文法属性番号を示しており，以後も同様である．

- 形態素解析による失敗も修正する場合

例文：お茶の木は中国から日本へ伝わった。

この例文の「お茶の木」を形態素解析にかけると次のようになる．

お茶 (b1100, #23043#-29945, #-29946, #-18430, #-26362) の木 (1100, #2357【類系】)

意味属性番号は #1 ~ #2716 までしか無く，上記の例ではその範囲外となっている．そこで次のように意味属性を振り直す．

お茶 (b1100, #858【茶】，#1931【休憩】，#1675【舞踊，演劇，諸芸】，#696【葉】，#1689【開会】，#675【樹木(その他)】) の木 (b1100, #673【樹木】，#772【材木】，#923【楽器】)

最期に，文意に合致する意味属性のみを選択する．

お茶 (b1100, #858【茶】) の木 (b1100, #673【樹木】)

- 複数の意味属性を残した名詞句

例文：少量のアルミを含有する．

形態素解析処理後：

少量 (#2597【多少】) のアルミ (b1100, #718【卑金属】，#710【元素】)

上記の例におけるアルミは，意味属性【卑金属】，【元素】でも文意に合致する．ゆえに上記の例のような場合は学習データとして，両者とも正しいので複数残す．

- 誤り

例文：ホテルのバーは午後 4 時に開きます．

形態素解析処理後：

ホテルのバー (b1000, #2595【単位】，#2567【多少】)

この例文における「バー」という名詞は，ホテルの飲み屋のような意味で使用されている．しかし「バー」に対して形態素解析が付与する【単位】【多少】は共に正しくない．従って上記のような例は学習データより除外する．同様に以下のような例も取り除く．

1. 色の深み (b1000, #461【土地】，1306【親近】)
2. オペレーティングシステムのプロンプト (b1100, #1835【奨励】)

## 2.6 規則の生成

2.5 節で作成したデータを用い規則を作成する。規則は、A・B1 対 1 の意味属性の組で作成する。この際、異なる英語表現で同じ組の意味属性があった場合は、学習データ中の頻度の高い英語表現を規則として登録する。また、手作業によるミスの影響を少なくするため、頻度が比較的多い英語表現に対しては、2 件以上ある意味属性の組を規則として使用する。

### 2.6.1 汎化の原理

ここで学習データの少なさによる正解率低下を防ぐために汎化による規則の生成を行う。汎化とは、意味属性体系における上位の意味属性の性質が下位の意味属性の性質に継承されるという特徴を利用し、規則に汎用性を持たせる処理である (文献 [2])。

### 2.6.2 汎化の効果

汎化の例として意味的に類似した関係にあり同じ英語表現を持つ「店【商店】のカレー【飯】」と、「レストラン【飲食店】のピザ【料理・その他】」を考える (図 2.4)。今「店のカレー」を学習データとした場合、汎化によって得られる規則は「【店舗等】の【料理】」である。この規則は、上下の包含関係より「レストランのピザ」にも適用される。つまり、一つの学習データから、汎用性を備えた規則が作成されたことになる。

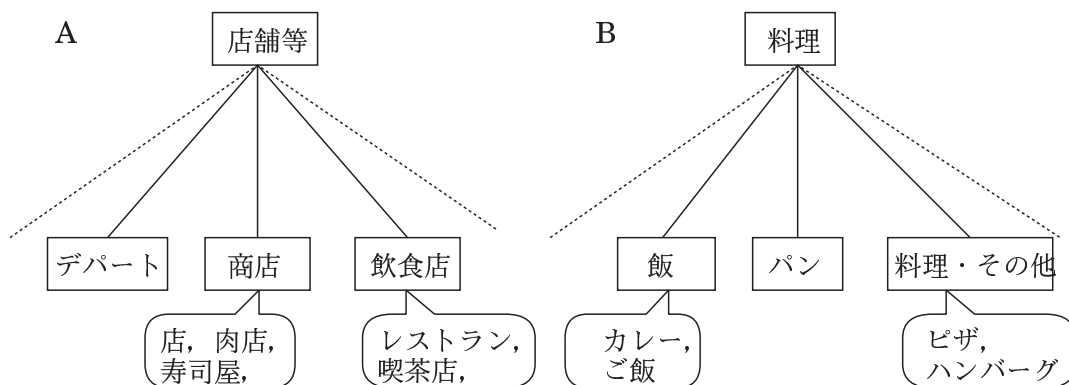


図 2.4: 両者の意味属性

### 2.6.3 汎化のアルゴリズム

汎化による規則の生成アルゴリズムを以下に示す。

1. 意味属性体系を，ボトムアップ式に，深さが深く意味属性が小さい方から順に，グループ化していく (図 2.5) . これにより意味属性体系を，計 789 グループにグループ化する .

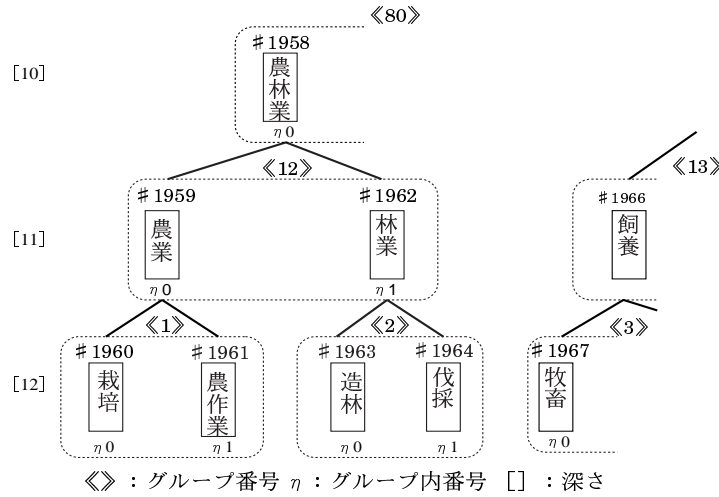


図 2.5: 意味属性体系のグループ化

2. グループ内の英語表現が 1 種類るとき上位の意味属性に英語表現と頻度を書き込む (図 2.6) .

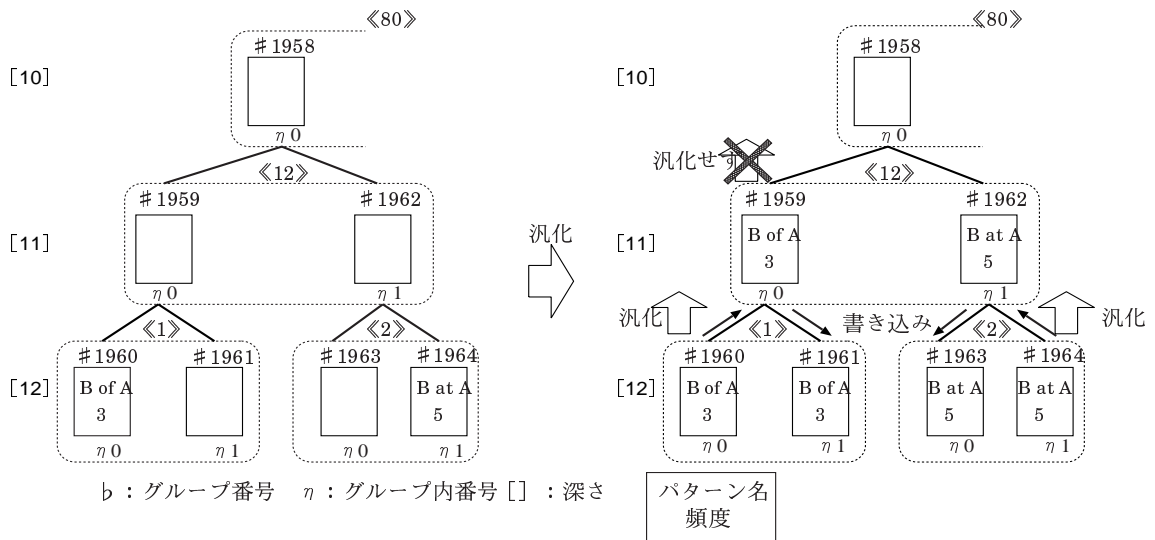


図 2.6: 汎化 1 次元

3. 2 の処理をグループ順に B, A 交互に行う. 2 次元の規則は図 2.7 のように汎化されていく. ここで, 図 2.7 が示す通り汎化の方向によって生成される規則が異なることがあるので場合によっては汎化方向を変更する必要がある.

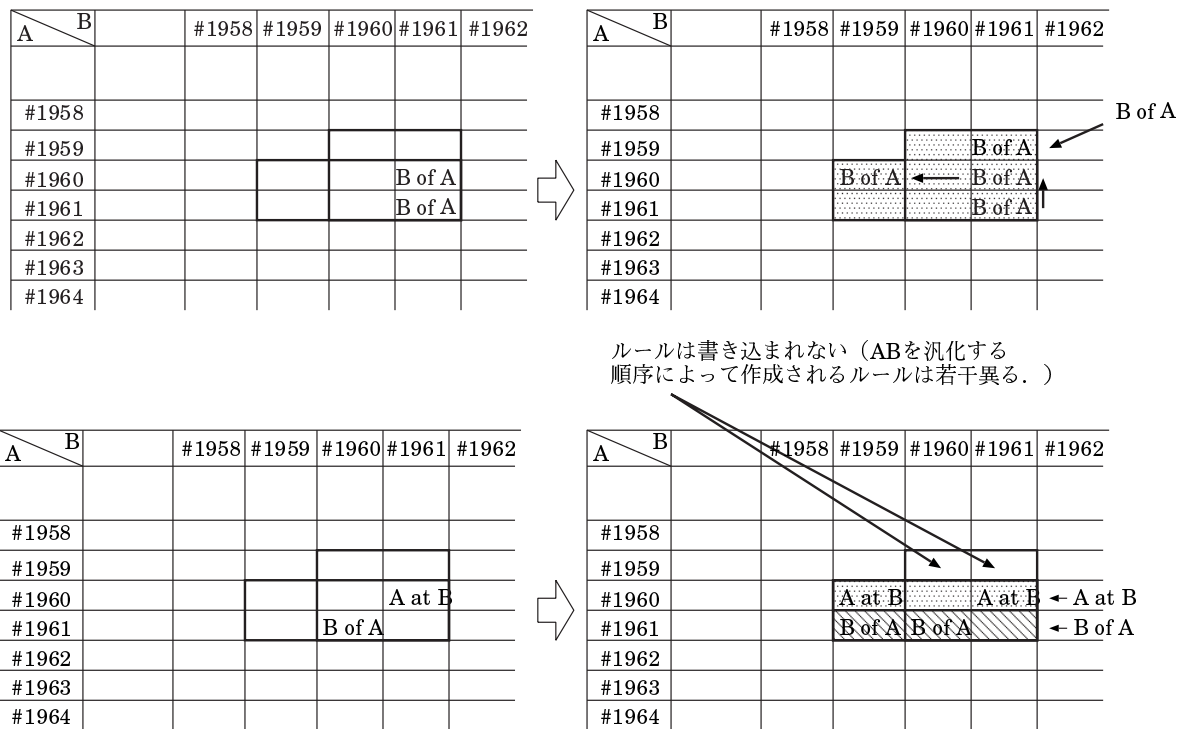


図 2.7: 汎化 2 次元

以上より作成した規則の件数と例をそれぞれ表 2.2 , 2.3 に示す .

表 2.2: 意味属性規則の作成数

英語表現	件数 (件)	英語表現	件数 (件)	英語表現	件数 (件)
A+B	46,315 件	形容詞 + B	46,188 件	B of A	152,876 件
A by B	1,560 件	B + A	20051 件	B on A	11,226 件
A' B	3,372 件	B about A	1,027 件	B to A	13,169 件
A for B	658 件	B against A	664 件	B with A	1,799 件
A from B	809 件	B as A	597 件	前置詞+A	4,420 件
A in B	1,504 件	B at A	5,082 件	所有格+B	10,448 件
A of B	6,107 件	B by A	2,573 件	A's B	41,175 件
A on B	1,010 件	B for A	17,894 件	単語	3,616 件
A to B	1,034 件	B from A	6,946 件	総数	440,545 件
A with B	201 件	B in A	23,894 件		

表 2.3: 意味属性規則

A 意味属性	B 意味属性	英語表現
#1680【スポーツ】	#1756【競争】	A + B
#125【友人】	#919【帳】	A' B
#220【神仏】	#1815【助力】	A for B
#220【神仏】	#1816【援助】	A for B
#220【神仏】	#1303【愛】	A for B
#841【野菜】	#2411【発酵】	A from B
#1147【情報】	#1031【概要】	A in B
#2336【力(物)】	#2592【度】	A in B
#496【川】	#2667【ふち】	A of B
#86【子】	#558【ほお】	A on B
#496【川】	#2446【基・源】	A to B
#906【桶・缶・樽】	#748【水】	A with B
#463【領土】	#238【学生】	形容詞+B
#2712【以前】	#2699【現在】	B A
#5【人間】	#1149【噂】	B about A
#2310【火】	#1402【用心】	B against A
#1826【規制】	#1936【従業】	B as A
#1234【宴会】	#1526【提唱】	B at A
#2667【ふち】	#729【石】	B by A
#360【選手】	#1209【失敗】	B by A
#988【乗り物(本体(移動(陸圏)))】	#1659【旅行】	B by A
#2421【怪我】	#800【薬品】	B for A
#862【たばこ】	#739【灰】	B from A
#463【領土】	#1471【景】	B in A
#901【冷暖房具】	#794【個体燃料】	B in A
#464【行政区画】	#2060【火事】	B in A
#408【図書館】	#902【出版物】	B in A
#388【場所】	#878【壁】	B of A
#1111【文章(全体)】	#1031【概要】	B of A
#819【衣服(本体(上半身))】	#821【衣服(部分)】	B of A
#1974【工事】	#2695【期間(自然・人間活動等)】	B of A
#990【乗り物(本体(移動(空圏)))】	#418【道路】	B on A
#2354【音】	#2453【影響】	B to A
#821【衣服(部分)】	#819【衣服(本体(上半身))】	B with A
#2655【冷暖房具】	#2655【周辺】	前置詞+B
#15【対称(単数)】	#1017【アイデア】	所有格+B
#95【兄】	#2406【呼吸】	A's + B
#2603【対】	#818【衣服(本体)】	単語

### 2.6.4 意味属性規則の適用について

以上の節より作成した意味属性規則を用いる場合，入力とするデータによっては，対応する意味属性規則が無い場合が存在する．学習データを増やすことにより，対応する意味属性規則のカバー率は増加するが，完全に対応させるためには相当数の規則が必要である．しかし，学習データは 2.5 節で述べた通り手作業で作成している．そのため，大規模な学習データを用意することは困難である．そこで意味属性規則の適用において，対応する意味属性規則が無い場合は以下の手順で規則を決定する．

1. 入力したデータに対する規則が無い場合，入力したデータに対して，上位の意味属性にあたる組の規則を検索して適用する．簡単のため，1 次元の例を図 2.8 に示す．図 2.8 では，3 つ上位の意味属性にあたる規則を検索する．検索数は，入力の意味属性の深さに最も近い英語表現を最大 3 件とする．
2. 複数の規則を検索した場合，学習データ時の頻度が最も多い規則を正解とする．

これにより，入力データに対応する規則が無い場合でも，入力データと意味的に近い関係にある規則を検索することが出来る．

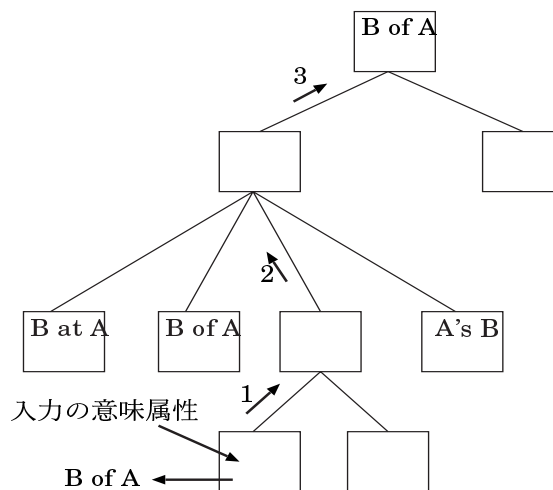


図 2.8: 上位意味属性の探索



## 2.7 相補的規則の作成

英語表現によっては意味属性より，名詞の字面や文の動詞に依存する．従って，意味属性の相補的な規則として字面・動詞規則を作成する．

### 2.7.1 字面規則の作成

「私の自転車」や「私の部屋」などの「所有格+B」の英語表現は「私の」という字面で英語表現が一意に決定できる．従って字面に依存度の高い以下の 6 種類については，翻訳規則として字面を追加記述する．

- A の字面について規則を作成した英語表現  
B at A，所有格 + B
- B の字面について規則を作成した英語表現  
B against A，B as A，前置詞 +A
- AB の字面を共に規則として記述した英語表現  
単語

登録単語は学習データ中に，出現頻度が 2 桁以上ある単語を採用する．登録数は，学習データにおける統計を基に，90 単語登録する．規則作成に使用した字面を 2.4 に示す．また字面規則の書式を図 2.9 に示す．図 2.9 は，左のカラムから順に

字面，英語表現番号，学習データ中の頻度  
となっており，規則の適用条件は字面の一致である．

#A名詞	genitive_b > 1	genitive_b is 27
彼	27	1639
私	27	864
あなた	27	458
彼女	27	421
我々	27	156
うち	27	131
貴社	27	106
僕	27	74
	:	

図 2.9: 字面規則の書式

### 2.7.2 動詞規則の作成

今回、翻訳対象とする英語表現を広げたため、文の用言に大きく依存する英語表現も多数でてくる。例えば「彼は上着の雪を払った。(He swept the snow from his coat.)」は、動詞「払う」によって英語表現が決定している。そこで、動詞に依存の高い6種類については、動詞の字面、共に共起する意味属性を規則として記述する。登録数は、学習データ 5000 件において、依存度が高いと思われる動詞を人手で選定し、合計 47 件の動詞を登録する。作成した規則を表 2.5 に示す。

また、作成した動詞規則の書式は図 2.9 になっており、左のカラムから順に、

動詞の字面、英語表現番号、用言文法属性、用言意味属性、共起する A 意味属性、共起する B 意味属性

となっており、後から容易に付け加えることができるようにしている。規則の適用条件は、動詞字面および共起する A・B 意味属性の一致である。

ところで、動詞規則については、用言文法属性・意味属性も記述しているが、現在は、動詞の字面及び、共起する A・B 意味属性を使用しようとしている。これは形態素解析の ALT-JAWS が、現在のところ用言意味属性の出力をできないためである。しかし、後にこの問題が改善されれば、より汎用性のある規則が作成できる見込みがある。ゆえに、用言意味属性及び用言文法属性も付与する。

#Verb	of	" A	by	B "	.	" A	by	B "	is	" 2 "	.
引く	2	2316	23	86, 54	594, 592						
持つ	2	2346	19	971, 972	2658						
知る	2	2386	30	238	556						
持ち運ぶ	2	2366	23	-1	957						
つかまえる	2	2416	23	537	821						
呼ぶ	2	2366	29	5	594, 592						
見覚える	2	2416	-1	238	555						
取る	2	2386	23	5	594, 592						
					:						

図 2.10: 動詞規則の書式

表 2.4: 字面規則

英語表現	作成した字面		件数 (件)	学習データにおける割合 (%)
所有格+B	彼	私	15 件	67.0%
	あなた	彼女		
	我々	うち		
	貴社	彼ら		
	僕	当社		
	私ども	貴殿		
	私たち	こちら		
	弊社			
B at A	学校	家庭	11 件	51.4%
	大学	店		
	会社	駅		
	レストラン 広場	オリピック		
前置詞+B	中	上	18 件	98.3%
	間	下		
	前	後		
	近く	方		
	陰	周り		
	周囲	向こう側		
	外側	間中		
	裏	後ろ		
	代わり	内側		
B against A	予防	用心	2 件	57.1%
B as A	経験	役	2 件	42.8%
単語	女の人	他の人	43 件	67.0%
	私のもの	畏敬の念		
	感謝の気持ち	残りの部分		
	世代の人々	先見の明		
	近所の人々	ドアの取っ手		
	丘の中腹	近所の人		
	よその人	象の鼻		
	ぼつきの柄	最悪の事態		
	農家の人	足の裏		
	血液の循環	時計の針		
	心配の種	羨望の的		
	柿の実	頭痛の種		
	年上の人	柿の木		
	他のもの	パンの耳		
	空の旅	机の上		
	自責の念	意思の疎通		
	鉛筆の芯	意見の一致		
	あなたのもの	アパートの管理人		
	柿の木	本の虫		
	結婚の申し込み	喉の渇き		
	座右の銘	木の皮		
	話の種			

表 2.5: 動詞規則

英語表現	動詞字面	「A」意味属性	「B」意味属性	登録件数 (件)
A by B	捕まえる	#537	#821	9 件
	引く	#54	#594 or #592	
	持つ	#971 or #972	#2658	
	知る	#238	#556	
	持ち運ぶ	*	#957	
	呼ぶ	#5	#594 or #592	
	見覚える	#238	#555	
	取る	#5	#594 or #592	
A for B	請う	#220	#1815 or #1816	3 件
	祈る	#220	#1303	
	応える	#125	#1815	
A from B	防ぐ	#677 or #841	#2411	4 件
	払う	#532	#737	
	救う	#5	#2079	
	許	#5	#1939	
A in B	突く	#5	#2459	9 件
	任せる	#865	#1779	
	入れる	#988	#967	
	攻撃する	#123	#2647	
	表示する	#1147	#1031	
	整える	#962	#1458	
	襲う	#123	#2648	
	変化する	#2336	#2592	
	保つ	#467 or #466 or #387	#2555	
A on B	触る	#537	#637	7 件
	叩く	#5	#558	
	キスする	#86	#558	
	打つ	#5	#558	
	マッサージする	#5	#582	
	立ち並ぶ	#418	#2645	
	まぶす	#543	#2569	
B from A	除く	#677 or #854	#702	16 件
	応じる	#238	#1511	
	従う	#971	#1824	
	集める	#385	#932	
	取り外す	#868	#1548	
	使う	#434	#1190	
	くぎる	#238	#889 or #423	
	借りる	#5	#1815 or #1816	
	割く	#1167	#2692	
	受ける	#48	#1699	
	答える	#136	#1511	
	纏める	#381	#1009	
	汚す	#439	#738	
	抜く	#908	#2497	
	買う	#5	#1316	
	吸収する	#757	#748 or #709	

\* ...どんな意味属性でもよい

## 第3章 実験

2章で作成した翻訳規則の精度を調べるため実験を行う。実験に使用する装置としてAT互換機(Linux: Athlon1.6GHz)を用い、プログラム言語はC++を用いる。実験は、以下に示す3種類の実験を行う。

- 実験1

実験内容 学習に使用しなかった標本よりランダムに抽出したデータ1000件を用い評価テスト

使用する翻訳規則 (1) 意味属性規則, (2) 意味属性規則+字面規則+動詞規則

狙い 生成した規則の評価

- 実験2

実験内容 すべての英語表現がほぼ等確率に含まれているオープンテストデータ267件での実験

使用する翻訳規則 (1) 意味属性規則, (2) 意味属性規則+字面規則, (3) 意味属性規則+字面規則+動詞規則

狙い 英語表現ごとの誤りの傾向の調査

- 実験3

実験内容 実験2のテストデータを用い意味属性を除く規則の評価実験

使用する翻訳規則 (1)A字面, (2)B字面, (3)AB字面, (4)単語

狙い 字面・動詞規則への依存度の調査

### 3.1 実験1

作成した規則の精度を評価するためオープンテストを行う。実験データは、収集した標本(2.3節参照)からランダムに1000件を抜き出したデータを用いる。また実験の際、実験データはあらかじめ形態素解析“ALT-JAWS”にかけておき、規則は、計算機に実装しておく。実験の評価基準は表3.1の通りとし、以下に評価基準の例を示す。

#### 評価基準の例

: 対訳と一致

例1: 雨戸の隙間

翻訳規則: B in A

対訳: B in A(a crack in the shutters)

例2: 雪の重み

翻訳規則: B of A

対訳: B of A(the weight of the snow)

: 対訳と一致していないが正解

例1: 患者の体温

翻訳規則: B of A(temperature of a patient)

対訳: A's B(the patient's temperature)

例2: ジャムの味

翻訳規則: B of A(taste of jam)

対訳: A B(the jam taste)

: 文によっては正解

例文1: 犬は泥棒の足にかみついた。

翻訳規則: B in A(The dog bit the thief in the leg.)

対訳: B by A(The dog got the thief by the leg.)

例文2: ぼくらはその男が女性のハンドバッグを盗むのを見つけた

翻訳規則: A's B(woman's purse)

対訳: B from A(We detected[caught] the man stealing a purse from a woman.)

×: 誤った英語表現

例文1: 大雪で列車の運行は混乱した

翻訳規則: A's B(train's service)

対訳: A B(A heavy snowfall threw train services into confusion.)

例文2: 羊の群れが仲良く草を食べている。

規則翻訳: A B(sheep flock)

対訳: B of A(Flocks of sheep have returned to their folds.)

尚評価では、翻訳後の英語表現のみを対象とし、訳語の適正は考慮しない。実験結果を表3.3に示す。

表 3.1: 評価基準

評価の種類	評価基準
	対訳と一致
	対訳と一致していないが正解
	文によっては正解
×	誤った英訳形式

表 3.2: オープンテストデータ 1000 文の内訳

英語表現	件数	英語表現	件数	英語表現	件数	英語表現	件数
A + B	100 件	A to B	5 件	B by A	11 件	前置詞 + B	79 件
A by B	10 件	A with B	4 件	B for A	41 件	所有格 + B	120 件
A' B	22 件	形容詞 + B	67 件	B from A	22 件	A's B	71 件
A for B	8 件	B + A	32 件	B in A	48 件	単語	29 件
A from B	5 件	B about A	8 件	B of A	152 件	総数	1000 件
A in B	14 件	B against A	6 件	B on A	35 件		
A of B	27 件	B as A	9 件	B to A	30 件		
A on B	7 件	B at A	29 件	B with A	9 件		

### 3.1.1 考察

3.1 節の実験において評価が × , となった名詞句について考察を行う．誤った規則に適用される名詞句は，誤る原因により分類することができる．誤る原因を以下に示す．尚考察中の「規則 1」は意味属性規則だけの規則，「規則 2」は意味属性規則と相補的規則 ( 字面規則 , 動詞規則 ) を組み合わせた規則を示している．

- × となる原因

1. 一つの意味属性だけでは，単語の性質を表しきれない名詞があるため失敗  
規則 1 : 33 件 / 268 件 ( 12.31% ) , 規則 2 : 30 件 / 215 件 ( 13.95% )

例文 1 : 屋上 の望遠鏡

対訳 : B on A ( the telescope on the roof )

規則 : B in A ( the telescope in the roof )

例文 2 : 戦後 の日本

対訳 : A B ( postwar Japan )

規則 : 形容詞 + B

今回意味属性の規則は，AB1 対 1 の意味属性の組として記述した．例文 1 における屋上は [ 上 ] と [ 家屋 ] の両方の意味を持つ．このような単語は，1 つの意味属性では性質を表しきれないため今回の規則においては誤った英語表現を返す．従って意味属性規則を，多数対多数の英語表現として記述する必要があると考えられる．

表 3.3: 実験1の結果

適用範囲	評価				
	規則				×
全体での 評価	意味属性	39.0%	16.6%	17.9%	26.5%
	意味属性+字面・動詞	49.9%	15.0%	12.8%	21.5%
形態素解析によるエラーを 除いた範囲での評価	意味属性	42.1%	18.0%	19.3%	20.6%
	意味属性+字面・動詞	53.3%	16.0%	13.7%	17.0%

...考察原因5, 6, 7を除くもの

2. 類似する意味関係にさまざまな英語表現が存在するため失敗  
規則1: 57件/268件 (21.27%), 規則2: 49件/215件 (22.79%)

例3: 列車の車掌とタクシーの運転手

例4: 俳優の演技と投手の技

例3の2つの名詞句は、意味的に類似しているがそれぞれ「conductors on this train(B on A)」と「taxi driver(A + B)」という英語表現を持つ。規則では、両者とも英語表現“ A + B ”であるため、「列車の車掌」は誤りとなる。この場合においては、意味属性での訳し分けは困難であると考えられる。

3. 学習データ不足

規則1: 52件/268件 (19.40%), 規則2: 44件/215件 (20.47%)

今回、標本中にデータ数が10件以上の英語表現を訳し分け対象としたが、そうした場合、下位の出現頻度が極めて低い英語表現は、学習データの不足という問題が発生する。また、出現頻度が多い英語表現についても学習データの作成に限度があるため、生成された規則がかなり汎化されている恐れがある。

例文5:

プール (#426【競技場】, #367【公共機関】, #1878【投資】) の水 #748【水】)

例文6:

エンジン (#964【原動機】, #991【乗り物(部分)】) の中心 (#2654【中心】, #2660【間】, #1030【要点】, #1029【主旨】, #2616【点(場所)】)

学習データが少ない場合上位検索による意味属性の適用(2.6.4章参照)が多く行われることになる。例文5では、この文における「プール」の相応しい意味属性は#426【競技場】である。しかし、学習データ不足だと、#367【公共機関】や#1878【投資】の文に合致しない意味属性で、規則の適用が数回行われる。これにより原文とは意味的に遠い関係にある規則が適用されることになり、誤った規則に適用される。



## 4. 一名詞句に多くの語意があるため失敗

規則1：50件/268件(18.65%)，規則2：28件/215件(13.02%)

例文7：

1. 仕事(#1936【仕事】)のあら(#2495【短所】，#799【屑・粕】，#2710【新旧・遅速】)

対訳：B with A(fault with work)

規則：B of A(fault of work)

2. 仕事(#1936【仕事】)のゴミ(#799【屑・粕】)

英訳：B of A(gabage of work)

規則：形容詞+B

3. 仕事(#1936【仕事】)の遅れ(#2127【停滞】，#2710【新旧・遅速】)

対訳：B of A(arrears of work)

規則：B of A(arrears of work)

例文7の場合，意味属性において，“仕事のあら”は，“仕事のゴミ”“仕事の遅さ”にもとれるため，正しい英語表現が決定できない．従って，これらの意味の絞り込みは結合価パターンとの併用や各要素の考慮など文レベルの解析が必要であると考えられる．他に

例：クラス(#2441【等級】#381【クラス】)の代表(#342【代理】#2439【正】#2435【類型】#119【成員】#2440【副】)

等の例が挙げられる．

ここで原因2, 3, 4は互いに依存し合う．従って，相互に考慮して改善していく必要がある．例えば，学習データを増やしても，“類似する意味関係にさまざまな英語表現が存在するために失敗”するケースが増える．

## 5. 形態素解析の際に失敗

規則1：32件/268件(11.94%)，規則2：25件/215件(11.63%)

このケースは，本研究の扱っている問題の範囲外なので除外する．

## 6. 文に合致する意味属性が付与されていない

規則1：31件/268件(11.57%)，規則2：28件/215件(13.02%)

例文8：

ワイン(#861【酒】)のコルク(#772【材木】)を抜いてください。

対訳：Please uncork the wine bottle.

例文9：

原爆の被害者は戦争という悲劇の生き証人(#344【本人】)だ

対訳：The atom bomb victims are living witnesses to the tragedy of war.

例文8において「コルク」は，対訳を見ても分かる通り「ワインの栓」という意味で用いられている．従って，ここで「コルク」に付与された【材木】は，適切ではない．

## 7. 固有名詞

規則 1 : 11 件/268 件 (4.10%) , 規則 2 : 10 件/215 件 (4.65%)

例文 10 : ワーズワース (#1【名詞】) の詩 (#1040【詩】)

規則 : 「所有格 + B」

例文 11 : シャガール (#1【名詞】) の絵 (#1048【絵画】 #938【像・書画】)

規則 : A's B

固有名詞は、一般名詞意味属性体系においては通常【名詞】として扱われ、一般名詞意味属性による訳し分けは、困難である。

## 8. その他

規則 1 : 11 件/268 件 (4.10%) , 規則 2 : 19 件/215 件 (8.84%)

他に、例文 12 に示す字面規則の誤った適用により失敗した名詞句や、意味属性体系における深さが浅すぎるために誤った場合などが挙げられる。

例文 12 :

トムは会社 (#374, #428) の宝 (#766, #2533) だ。

対訳 : Tom is a treasure to our company.

規則 : treasure at company (B at A : 字面規則)

現状では、上記の例のようなものはあまり誤りの比率を占めな。しかし、今後字面規則等を増やす場合には、単独の名詞のみで英語表現を決定しない方が良いように思われる。

- (文によっては正解) となる原因

## 1. 動詞に依存

規則 1 : 57 件/179 件 (31.84%) , 規則 2 : 55 件/128 件 (42.97%)

例文 13 :

彼の指示 を受けるべきだ。

対訳 : You must receive instruction from him.

規則 : 所有格 + B(his instruction)

例文 14 :

人の危難 を救う

対訳 : to deliver man from danger

規則 : A's B(one's danger)

例文 13 において、名詞句「彼の指示」のみに着目すると、翻訳は“ his instruction ”となるが、「受ける」という用言があるため、「彼からの指示」という意味合いを持ち、この文においては誤った英語表現となる。

## 2. 文の意味や構造に依存

規則 1 : 63 件/179 件 (35.20%) , 規則 2 : 44 件/128 件 (34.38%)

例文 15 :

本屋 (#434【商店】 , #273【商人】) の商売は夏は全く暇です

対訳 : The bookseller's trade is at a standstill in summer.

規則 : trade of bookstore

例文 15 において、本屋が、人 (#273【商人】) を指していれば「A's B」であり、本屋が、(#434【商店】) そのものを指していれば「B of A」である。こ

のため，文脈を考慮しなければ一意に英語表現を決定できない．  
他に，

例文 16：晩 (#2686【夕】 #687【夜】 #2708【終り】) の食事

- (a) evening meal
- (b) dinner this evening

のように，文脈により「晩の」が「今晚の」という意味で用いられると英語表現が異なる．

### 3. 絞り込みに失敗

規則 1：51 件/179 件 (28.49%)，規則 2：23 件/128 件 (17.97%)

例文 17：

彼らの宿舎には市内のホテルが使われている

対訳：They are staying at hotels in the city.

規則の出力：

(#464,#374) 頻度:1 A B

(#464,#437) 頻度:1 B in A

例文 18：

彼は生命の起源はどうだったのかといろいろ考えた

対訳：He wondered a lot about the origin of life.

規則の出力：

(#2532,#2445) 頻度:1 A's B

(#2385,#2445) 頻度:1 B of A

例文 17 における「市内 (#2623【内部】，#464【行政区画】) のホテル (#437【宿泊施設】 #374【企業】)」は，その意味属性の組み合わせによって

#464【行政区画】 #374【企業】 A B

#464【行政区画】 #437【宿泊施設】 B in A

という翻訳規則が存在するため一意に英語表現が絞り込めない．語意数が多い名詞は，必然的にこのケースになりやすい．

## 3.2 実験2

英語表現ごとの誤りの傾向を調べるため、翻訳対象としている全ての英語表現が、ほぼ等確率で含まれているオープンデータ 267 件を用いた実験を行う。実験データは英語表現 1 つにつき 10 件含まれるようにしているが、標本数などの理由により英語表現によっては 10 件を下回る。また評価は、

○：対訳と一致

×：対訳と不一致

の 2 種類で行う。

閾値制御について

ところで、動詞規則及び字面規則を作成した英語表現については、意味属性規則への依存度が低くなるものと思われる。従って、動詞・字面規則を作成した英語表現については、汎化する際の閾値を高くし、意味属性規則における影響を少なくした方が全体の翻訳精度が向上すると思われる。その際の閾値は、「学習データにおける割合」(表 2.4 参照) 及び「動詞規則の数」(表 2.5 参照) に基づき設定する。

実験は、これをふまえ閾値制御を行った意味属性規則を用いた実験も同時に行う。

### 3.2.1 考察

結果を表 3.3 に示す。表中の評価は件数でつけており、p1 ~ p4 は、意味属性規則のみを使用した評価との相対で表している。のパーセンテージは、p1 : 31.85%、p2 : 37.90%、p3 : 39.11%、p4 : 40.32%である。

- 動詞規則及び字面規則を追加、共起意味属性規則の閾値制御を行うことにより精度の向上が見られる。これを実験 1 に適用した場合、3%ほど向上する。
- 単語については、辞書を直接参照することによりかなり精度が良くなる。ただ、単語のほとんどについては他の言い回しが存在する。
- 名詞に形容詞型や副詞型が存在するために×や○になるものが多数ある。形容詞に関しては、単語と同様辞書に形容詞型があるものは、形容詞型を優先的に用いた方がよいと思われる。また、副詞に関しては、文の構造を考慮し、用いる必要があるように思われる。
- A、B が位置関係にあるものは、名詞句単独では判断が難しい。

例：船の上

“ 船の甲板 ”か“ 船の上空 ”かによって、訳語も英語表現も異ってくる。このような、位置関係について研究したのものもある(文献 [6])。従って、A、B が位置関係にあるものは更に文献 [6] を参考にさらに細かな翻訳規則が必要である。

- 字面規則は、全体に対する統計情報を基に作成したためかなり不足部分、及び適切でないものが含まれる。“ 次の ”、“ 他の ”などのものはその字面によって“ next ”、“ other ”と決定することができるが、これらの字面は、全体に対する頻度の比率が少ないため、規則に加えていない。従って、字面規則を記述する名詞の選択方法を、新たに考える必要がある。

表 3.4: 実験2の結果

英語表現									×				E <sup>5</sup>	総数
	p1 <sup>1</sup>	p2 <sup>2</sup>	p3 <sup>3</sup>	p4 <sup>4</sup>	p1	p2	p3	p4	p1	p2	p3	p4		
A + B	4	-	-	+1	1	-	-	-	5	-	-	-1	0	10
A by B	3	-	-	-	0	-	-	-	6	-	-	-	0	9
A' B	2	-	-	-	0	-	-	-	8	-	-	-	0	10
A for B	1	-	-	+1	0	-	-	+1	7	-	-	-2	1	10
A from B	1	-	+1	-	0	-	-	-	5	-	-1	-	0	6
A in B	2	-	-	-	0	-	-	-	8	-	-	-	0	10
A of B	6	-	-	+1	1	-	-	-	1	-	-	-1	1	9
A on B	3	-	+1	-	0	-	-	-	4	-	-1	-	1	8
A to B	1	-	-	-	0	-	-	-	3	-	-	-	2	6
A with B	1	-	-	-	0	-	-	-	4	-	-	-	0	5
形容詞 + B	5	-	-	-	1	-	-	-	2	-	-	-	0	8
B + A	3	-	-	-	1	-	-	-	4	-	-	-	0	8
B about A	2	-	-	-	0	-	-	-	7	-	-	-	1	10
B against A	4	+1	+1	+1	0	-	-	-	3	-1	-1	-1	0	7
B as A	1	+1	+1	+1	0	-	-	-	8	-1	-1	-1	0	9
B at a	1	+2	+2	+2	0	-	-	-	8	-2	-2	-2	2	9
B by A	2	-1	-1	-	0	-	-	-	8	+1	+1	-	0	10
B for A	3	-1	-1	-1	1	-1	-1	-1	4	+2	-2	-2	2	10
B from A	1	-	+1	+1	1	-	-1	-	8	-	-1	-1	1	11
B in A	2	-	-	-	2	-	-	-	6	-	-	-	0	10
B of A	5	-2	-2	-	1	-	-	-	4	+2	+2	-	0	10
B on A	6	-	-	-	0	-	-	-	4	-	-	-	0	10
B to A	2	-	-	-	1	-	-	-	6	-	-	-	0	9
B with A	3	-	-	-	0	-	-	-	5	-	-	-	0	8
前置詞 + B	2	+8	+8	+8	1	-1	-1	-1	7	-7	-7	-7	0	10
所有格 + B	5	+4	+4	+4	4	-4	-4	-4	1	-	-	-	0	10
A's B	3	-	-	-2	2	-	-	-2	0	-	-	-4	7	12
単語	2	+1	+1	+1	2	-1	-1	-1	6	-	-	-	0	10
総数	79	94	97	100	19	12	11	7	146	137	135	135	19	267

<sup>1</sup> p1...共起意味属性規則

<sup>2</sup> p2...共起意味属性規則+字面規則

<sup>3</sup> p3...共起意味属性規則+字面規則+動詞規則

<sup>4</sup> p4...閾値制御を行った共起意味属性規則+字面規則+動詞規則

<sup>5</sup> E...形態素解析エラー

### 3.2.2 意味属性規則の分布について

実験2の検討を進める上でいくつかの英語表現が「A's B」等の特定の英語表現に間違えるケースがあった。そこで、作成した意味属性規則の配列の分布がどのようになっているかを視覚的に捉えるため図3.1に英語表現ごとに、規則の意味属性の分布を一部載せる。

図3.1から分かることは、英語表現によっては意味属性に殆んど偏りが見られず、意味属性の範囲全般に規則が分布している事である。図3.1から、英語表現によっては意味属性だけの翻訳は、あまり効果が得られないように見受けられる。これを解決するために、分散が高い英語表現について閾値制御や入力制御を何通りか試してみたが大きな効果は得られない。このことから意味属性のみを用いた規則では、劇的に翻訳精度を向上させるのは困難に思われる。

この解決策として、意味属性規則を用いる前に、文法属性や、「has-a」/「is-a」関係等を用いて、似通った意味属性の分布を示すパターンを分離することが挙げられる。

#### 文法属性

形態素解析プログラムALT-JAWS[8]では、文法属性として品詞を、26種類に分類している。今回の学習データでは、統計情報を基に規則を作成しようとした場合、すべての英語表現において「一般名詞+一般名詞」の組合せが多いが、文献[3]などによるアプローチの方法をとることにより、有効に利用できると思われる。

#### 「has-a」/「is-a」関係

意味属性体系[1]は、上位-下位関係の「is-a」と、全体-部分関係の「has-a」関係が存在する。「AのB」名詞句においては、「has-a」関係となるものが比較的少ないため大きな要因にはならないかも知れないが、機械規則として容易に規則化できるメリットがある。

また意味属性規則において、A・Bの意味属性のどちらか一方に偏った傾向が見られればそれは、1つの意味属性だけを使用した方が精度が向上すると思われる。

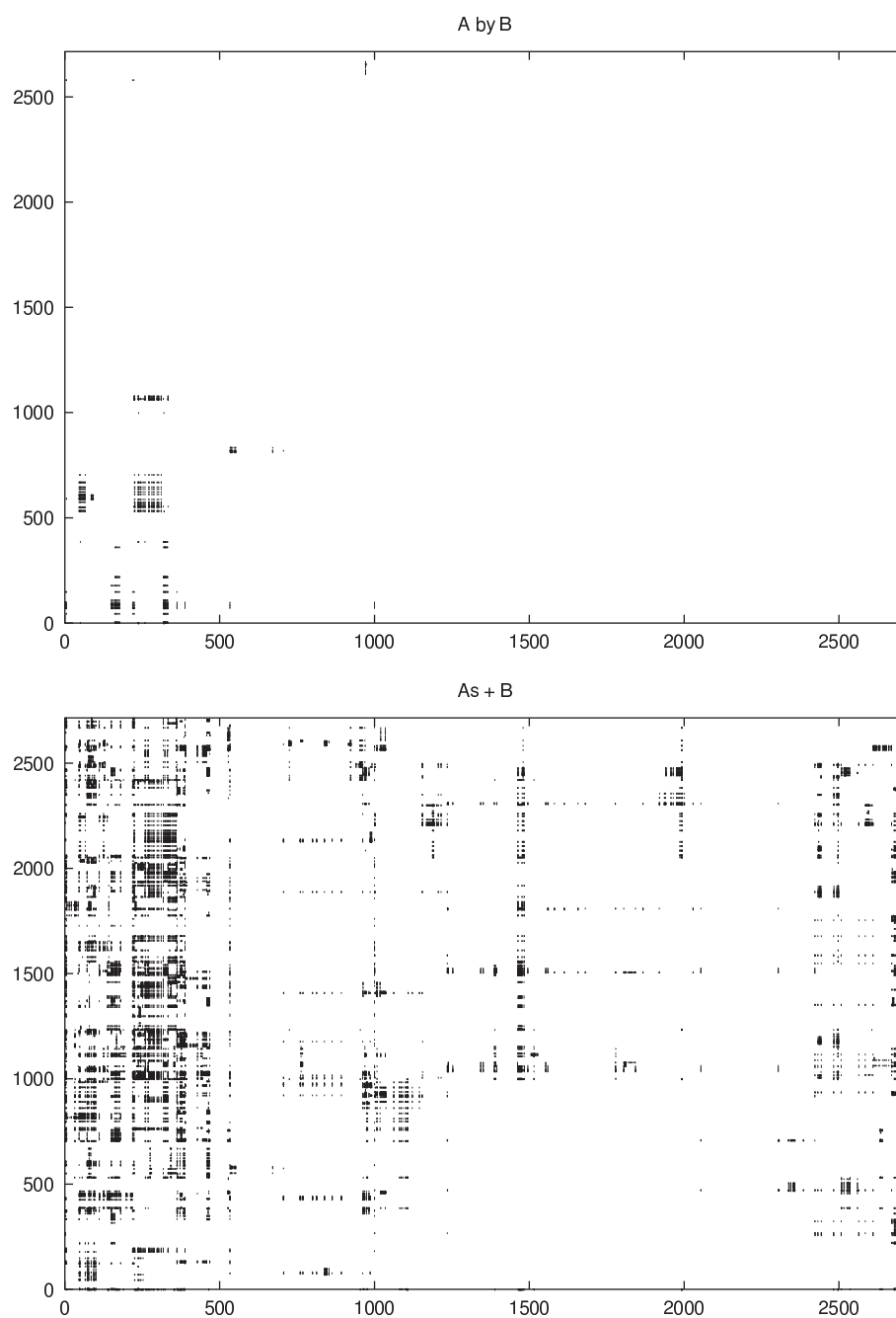


図 3.1: 意味属性規則の分布図

### 3.3 実験3

3.1節の実験において意味属性規則に字面・動詞規則を効率良く作成し組み合わせることで約70%の精度が得られることが分かった。ここでは、意味属性規則に対する動詞・字面規則の依存度を調べるため字面・動詞規則のみを用いた実験を行う。尚、実験データは3.2節において用いたオープンデータを用いる。結果を表3.5に示す。評価は、3.2節と同様

○：対訳と一致  
 ×：対訳と不一致  
 の2種類のみで行う。

表 3.5: 実験3の結果

規則 評価	A 字面規則	B 字面規則	AB 字面規則	動詞規則
	6.85%(17件)	6.45%(16件)	0.81%(2件)	4.03%(10件)
×	6.04%(15件)	1.21%(3件)	1.21%(3件)	1.21%(3件)

表3.5より、今回作成した字面・動詞規則全体では18%ほどになる。動詞規則は、標本全体における出現頻度が低い英語表現に対して有効であった。ことから意味属性による規則作成は比較的出現頻度が高いものに絞り、動詞規則と効果的な併用をした方がよいと考えられる。



## 第 4 章 検証

4 章では、3 章において正解率低下の原因になりうる要素について詳しく検証していく。検証項目は、

1. 学習データの数
2. 意味属性規則作成時における汎化、及びその規則の適用順序

の 2 点である。

### 4.1 学習データの検証

今回の学習データは、5000 件 (2.5.1 節参照) 用いていた。しかし、実際学習データの整備がどの程度必要か分からない。そこで、学習に使用していない標本から 1000 件ランダムに取り出して翻訳精度の評価を行う。尚、今回の評価は、

：対訳と一致

：計算機が返す英語表現の候補中に対訳と同じものがある

×：候補中に対訳にも同じものが存在しない

の 3 つに分けて評価した（“間違っているでもこの表現でも可”というものは今回いれていない）。

表 4.1: 学習データ検証結果

学習数		1000 件	2000 件	3000 件	5000 件
汎化なし		4.94%	8.88%	9.59%	10.82%
		0.00%	0.10%	0.10%	0.31%
	×	95.10%	91.02%	90.31%	88.88%
汎化あり		41.22%	38.78%	38.57%	40.92%
		10.41%	13.88%	13.17%	13.16%
	×	48.37%	47.35%	48.16%	45.92%

形態素エラー 20

この結果より、汎化を行った後の規則は 2000 件以上になれば となる数はあまり変化しないことが分かる。しかし 3.1 節の実験より、上位意味属性の検索による意味属性の適用が全体の 2 割ぐらい締めていた。ゆえに、今後さらに学習データを増やす必要があるように思われる。

## 4.2 汎化・適用順序の検証

次に、汎化・適用順序の検証を行う。検証は、以下に示す項目について検証を行う。

1. 一つの英語表現が汎化できる回数の上限を設定
2. 汎化の順序を B→A から A→B に変更

### 4.2.1 汎化回数の上限

これまででは一つの英語表現について汎化する回数の上限を設けていなかったが汎化を上限もなく過度にしてもよいかを検証するため汎化回数を 1~5(図 4.1) に変更してその精度の変化について検証する。検証には実験 2 で用いたデータ 267 件を用いる。結果を表 4.2 に示す。

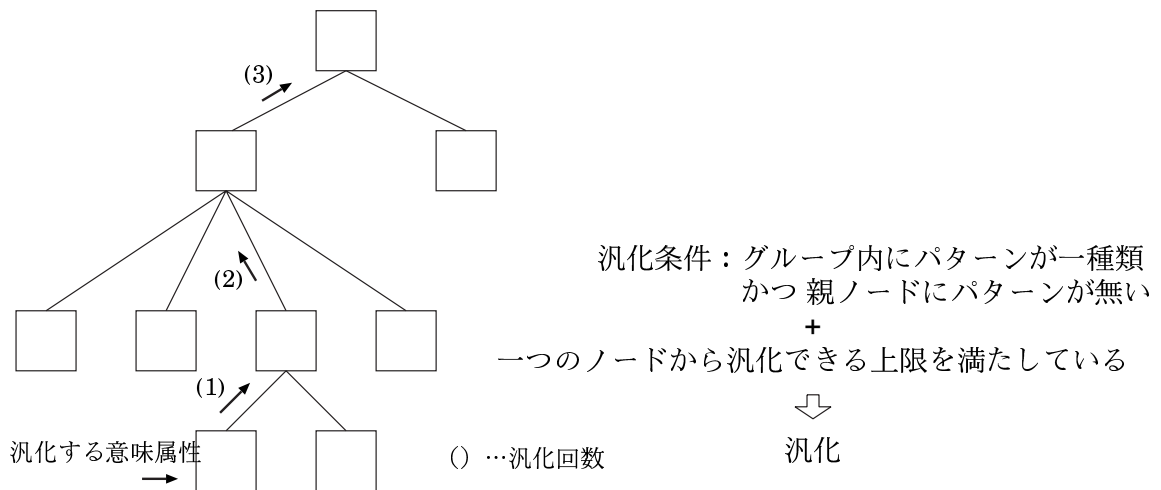


図 4.1: 汎化の上限

表 4.2: 汎化回数の変化に伴う精度の遷移

汎化回数 \ 評価	1	2	3	4	5	制限無し
	69 件	73 件	76 件	80 件	77 件	79 件
	9 件	14 件	18 件	17 件	19 件	19 件

この結果より汎化の上限は、特に限度を設けなくてもよいことが分かる。ただ、汎化回数を制限しすぎると明らかに正解率が悪くなるので、下限は 3 回ぐらいがよいと思われる。

### 4.2.2 汎化順序の変更

今回、汎化の順序を B を起点とした汎化、次に A を起点とした汎化、という繰り返し順序を取っている。これは名詞句「A の B」においては B の方に意味の重み（「庭の噴水」なら噴水がメイン）を置くことが多いためである。ここでは、汎化順序の妥当性を調べるため、汎化順序を A→B とした実験を行う。実験データとして 4.2.1 節と同じ 267 件のデータを用いる。

表 4.3: 汎化順序の変更に伴う正解率の変化

評価	汎化順序	
	A→B	B→A
	76 件	78 件
	20 件	19 件

結果を見ると、多少ではあるが B→A の方が精度がよくなっている。従って、汎化の順序は B→A が妥当であるといえる。

## 第5章 結論

### 5.1 おわりに

本研究では、「AのB」型名詞句の対訳付き標本を約3万件収集し、その標本中の英語表現28種類について、意味属性の共起パターンを用いて翻訳規則を作成した。また、意味属性による翻訳規則の相補的な規則として、字面・動詞による規則も作成した。計算機による実証を行ったところ、意味属性のみの規則による翻訳精度は約55%であり、字面・動詞規則を併用した翻訳規則の精度は約65%であった。しかし、形態素解析による失敗を除くと正解率はそれぞれ約59%、約70%であった。他に、文脈によっては、正解となるものが19%、20%と比較的多く、明確な誤りとなったものが21%、17%となることから、よい精度の翻訳規則を作成できたといえる。

### 5.2 今後の課題

今後の課題としては、

1. 文レベル(格要素等)での解析
2. 訳語も考慮した翻訳規則
3. 多段的な翻訳
4. 文献[3]によるアプローチ

が挙げられる。

1. 現在の翻訳規則は、動詞規則を多少考慮したとは言え、格要素など文全体の構造や、意味を考慮したものではない。従って、結合価文法との併用した翻訳方法の作成が必要である。結合価パターンとの併用の例を以下に示す。

結合価パターン

例文：N1がN2のN3を締める

N1 screw N3 on N2

[N2(904 容器)N3(“蓋”)]

この例では、上記の結合価パターンが用いられる場合、「N2のN3」は「B on A」に翻訳すれば良いことが分かる。また、格要素についての例を示す。

格要素

例文：馬のたてがみ

I 日本文：女の子はこわごわ馬のたてがみに触った。

英文対訳：The little girl touched the horse on the mane timidly.

II 日本文：馬のたてがみは素晴らしい。

英文対訳：A horse's mane is beautiful.

「馬のたてがみ」が、“ が ”のように主格として用いられているのか、“ を ”のように目的格として用いられているかによって英語表現が異なる。

2. 本研究では、「A の B」型名詞句の英語表現のみを考慮し、訳語は考慮していない。この翻訳方法を実用化段階に持って行くためには、訳語も同時に翻訳する必要がある。

3. 今回の研究では、意味属性を主とする規則で「A の B」型名詞句を 28 の英語表現に翻訳することを試みた。しかし、意味属性のみでは、約 6 割程度の翻訳しかすることができない。従って、他の文法属性などの要素で英語表現を 2 分割できれば、さらに翻訳精度が向上すると思われる。

4. 今回作成した翻訳規則の精度は、文献 [3] の精度に比べ劣る。これは、対象とする英語表現を増やしたことが影響するが、さらに個々の英語表現の翻訳精度を向上させるためには、文献 [3] と同じ、人による規則作成のアプローチが必要と思われる。

ところで、今回英語表現を増やすため、2.3 節に示した標本の収集方法をとった。このため、文献 [3] より、英語表現の数が少ない。(文献 [3] では、英語表現を 57 種類に分類している)。従って、まず文献 [3] と英語表現において、整合性を保つことが必要である。

## 参考文献

- [1] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦, “日本語語彙体系”, 岩波書店, 1997.
- [2] 池原悟, 中井慎司, 村上仁一, “日本語名詞句に対する意味的構造規則の自動獲得”, 第4回 LACE 研究会, 1999.
- [3] 宮本健司, 池原悟, 村上仁一, “意味関係に着目した「AのB」型名詞句の翻訳規則”, 卒業論文, 鳥取大学工学部知能情報工学科, 2000.
- [4] 守谷有司, 宮本健司, 池原悟, 村上仁一, “意味関係に着目した「AのB」型名詞句の翻訳規則”, 情報処理学会大63回全国大会, pp.269-270, 2001.
- [5] 横山昌一, 加藤貴子, 廣重拓司, “格助詞「の」の分類と解析”, 情報処理学会第1回年次大会, pp.137-140, 1995.
- [6] 廣瀬文男, 池原悟, 村上仁一, “日本語の位置的関係表現の意味的対応関係”, 平成12年度自然言語処理学会年次大会, A6-2, pp.490-493, 2001.
- [7] 村上仁一, 池原悟, 徳久雅人, “日本語英語の文対応の対訳データベースの作成”, 第7回 LACE 研究会, 2002.
- [8] 形態素解析プログラム ALT-JAWS, alt@nlp.ntt-at.co.jp
- [9] 島津明, 内藤昭三, 野村浩郷, “助詞「が」が結ぶ名詞の意味関係の subcategorization”, 情報処理学会言語処理研究会, 86-NL-53, pp.1-8, 1936.
- [10] 池原悟, 中井慎司, 村上仁一, “多義解消のための構造規則の生成方法と日本語名詞句への適用”, 自然言語処理論文誌, Vol.6, No.1, pp.237-266, 1986.