

## 概要

ブログ記事から旅行時の行動を分析する研究が、行われている。その中に、「どこに何があるか」を収集するタスクがある。なぜならば、場所とそこにある存在物がわかるだけでも観光分析にとって有益だからである。

用言パターン [1] を用いると、1つの文から場所と存在物を解析できる。しかし、ブログでは2文に分けて、場所と存在物を述べることがあるという問題がある。2文の内、1文目で解析できる情報と2文目で解析できる情報を単にまとめてしまえば、誤りを含むからである。そこで、本研究では、格情報および動詞の情報を持つパターンを対で用いて、場所と存在物の解析を行う手法を提案する。

具体的には、まず、ブログ文から連続する2つの動詞述語文の対を複数抽出し、場所と存在物を含むような対から、1文目のパターン (FP) と2文目のパターン (SP) で構成される、場所と存在物抽出用のパターン対を作成する。作成したパターン対 59 ペアでパターン辞書を構築し、日本語語彙大系における用言意味属性を利用して、パターン対の用言部分の網羅性を拡張する。

パターン辞書を用いた手法を評価するため、テスト用の2文対 47 件に人手で正解データを作成する。答えが出るべきか・出ないべきかに着目し、テスト用の2文対 47 件の内、正解が出るべき2文対 26 件にはそれぞれ1つ以上の正解データ (計 35 個) を付与し、正解が出ないべき2文対 21 件には空のデータを正解として付与する。次に、テスト用の2文対を対象に、3つの手法で抽出実験を行う。1つ目の手法は、構築したパターン辞書を用いて場所と存在物を抽出する手法である。2つ目の手法は、パターン対を分解し、FP と SP の全ての組み合わせで再構成したパターン辞書を用いて場所と存在物を抽出する手法である。3つ目の手法は、意味属性を利用して場所と存在物の全ての組み合わせを得る手法である。

3つの手法の実験結果について、抽出結果と正解データの  $F$  値で手法の評価を行う。

抽出実験の結果、手法1では、 $F$  値が 0.54 となった。手法2では、0.65 となった。手法3では、0.58 となった。手法2でもっとも  $F$  値が高くなっており、パターン辞書を用いない手法よりわずかに精度が向上した。今後の課題は、パターン辞書の拡充と見直し

により，抽出精度をより向上させることである．

# 目次

第1章	はじめに	1
第2章	場所と存在物の解析	2
2.1	1文からの人手による場所と存在物の解析	2
2.2	2文からの人手による場所と存在物の解析	2
2.3	1文からの機械的な場所と存在物の解析	2
2.4	問題設定	3
第3章	パターン辞書の構築	4
3.1	パターン辞書構築の概要	4
3.2	構築方法	4
3.2.1	手順	4
3.2.2	用いるタグセット	7
3.2.3	パターン辞書の拡張	7
3.3	パターン辞書の構築結果	7
3.3.1	パターン対の作成	7
3.3.2	用言意味属性を利用したパターン辞書の拡張	8
第4章	パターン辞書による場所と存在物の抽出	10
4.1	場所と存在物の組の抽出手順	10
4.2	パターン対による情報抽出	10
4.2.1	パターン照合	10
4.2.2	patlap用パターン対	12
第5章	実装	13
5.1	システムの概要	13
5.1.1	手順1から手順5：パターン辞書の構築	13

5.1.2	手順6：パターン辞書の形式変換	13
5.1.3	手順7から手順9：2文からの場所と存在物の組の抽出	13
<b>第6章</b>	<b>2文からの場所と存在物の抽出実験</b>	<b>18</b>
6.1	実験の目的	18
6.2	テスト文	18
6.3	正解データ	18
6.4	実験方法	19
6.5	評価方法	19
6.5.1	評価値	19
6.6	実験結果	20
6.6.1	抽出例	20
6.6.2	誤り例1	21
6.6.3	誤り例2	22
<b>第7章</b>	<b>考察</b>	<b>24</b>
7.1	手法1の性能が低い問題について	24
7.2	動詞の置き換えによるパターン辞書の拡張について	24
7.3	手法2のパターン対の再構成について	24
<b>第8章</b>	<b>おわりに</b>	<b>25</b>

# 目 次

2.1	用言パターンの例 . . . . .	3
3.1	パターン辞書構築のフロー . . . . .	5
3.2	CaboCha の出力例 . . . . .	6
3.3	タグ付き単文対の例 . . . . .	8
3.4	パターン対の例 . . . . .	8
3.5	動詞置き換えの例 . . . . .	9
4.1	場所と存在物の組の抽出のフロー . . . . .	11
4.2	用言パターン対の例 . . . . .	11
4.3	patlap 向けのパターン形式 . . . . .	12
4.4	パターン対の例 . . . . .	12
5.1	システムの実装 . . . . .	14
5.2	入力文の例 . . . . .	15
5.3	SPM の出力の例 . . . . .	16
5.4	出力の例 . . . . .	17
6.1	テスト文の例 3 . . . . .	21
6.2	手法 1, 手法 2 において適合したパターン 1 . . . . .	21
6.3	テスト文の例 1 . . . . .	22
6.4	手法 1, 手法 2 において適合したパターン 2 . . . . .	22
6.5	テスト文の例 2 . . . . .	23
6.6	手法 1, 手法 2 において適合したパターン 3 . . . . .	23

# 表 目 次

6.1	6.3 節の正解データを場所と存在物の抽出結果 . . . . .	20
6.2	$\phi$ のペアを正解として付与しなかった場合の場所と存在物の抽出結果 . .	20
6.3	出力された場所と存在物の組 1 . . . . .	21
6.4	出力された場所と存在物の組 2 . . . . .	22
6.5	出力された場所と存在物の組 3 . . . . .	23

# 第1章 はじめに

ブログ記事から旅行時の行動を分析する研究が，行われている．その中に「どこに何があるか」を収集するタスクがある．なぜならば，場所とそこにある存在物がわかるだけでも観光分析にとって有益だからである．

用言パターン [1] を用いると，1つの文から場所と存在物を解析できる．しかし，ブログでは2文に分けて，場所と存在物を述べることがあるという問題がある．2文の内，1文目で解析できる情報と2文目で解析できる情報を単にまとめてしまつては，誤りを含むからである．

そこで，本研究では，各および動詞の情報を持つパターンを対で用いて，場所と存在物の解析を行う手法を提案する．

第2章では，場所と存在物の解析方法について述べる．第3章では，パターン辞書の構築について述べる．第4章では，パターン照合について述べる．第5章では，システムの実装について述べる．第6章では，ブログ文から得たテスト文を用いて，実験を行い，比較手法と比較する．第7章では，考察を行う．第8章では，まとめを行う．

## 第2章 場所と存在物の解析

本章では、ブログ文からの人手による情報解析の方法、および、ブログ文からの機械的な情報解析の問題について説明する。

### 2.1 1文からの人手による場所と存在物の解析

場所と存在物は、動詞と格の繋がりを解析することで認識できる。例えば、「門司でフク井を食べました。」には、「門司で食べる」と「フク井を食べる」という2つの情報が含まれており、「門司にフク井が存在する」という情報が解析できる。

### 2.2 2文からの人手による場所と存在物の解析

1文からの場所と存在物の解析と同様の方法に加え、話の流れを推測することで、2文から場所と存在物を認識できる。例えば、「門司までドライブに行ってきました。フク井を食べてきました。」という2文からは、それぞれ場所「門司」と存在物「フク井」が解析できる。2文の格情報、および、「行く」と「食べる」という動詞の繋がりにから、「行った先」で「食べた」ことを推測し、「行った先」に「食べた物」が存在すると認識する。つまり、「門司でフク井を食べた」ことを推測し、「門司」に「フク井」が存在することを認識する。この方法で「門司にフク井が存在する」という情報が解析できる。

逆に、「フク井を食べてきました。その後、香住に行きました。」という2文からは、それぞれ存在物「フク井」と場所「香住」が解析できる。しかし、2文の格情報、および、「食べる」と「行く」という動詞の繋がりにからは、「香住にフク井が存在する」ことは解析できない。

### 2.3 1文からの機械的な場所と存在物の解析

1文からの場所と存在物の解析は、図2.1のようなパターンを用いて、機械的に解析が行える。意味属性制約は、パターン内の変数がとる名詞の範囲である。



- 
- (1) 用言パターン： $N1$  で食べる  
意味属性制約：[ $N1$  (場所)]
- (2) 用言パターン： $N1$  を食べる  
意味属性制約：[ $N1$  (具体物)]
- 

図 2.1 用言パターンの例

例えば、「門司でフク井を食べる。」という入力文の場合、2つのパターン「 $N1$  で食べる」、「 $N1$  を食べる」のそれぞれと適合する。1つ目のパターンで場所「門司」を、2つ目のパターンで存在物「フク井」を獲得できる。これにより、機械的に「門司にフク井が存在する」ことが解析できる。

## 2.4 問題設定

2文からの場所と存在物の解析は、機械的に行えない。例えば、「門司に行った。フク井を食べた。」という入力文の場合、パターンを用いると単文の場合と同様に、1文目からは場所「門司」が、2文目からは存在物「フク井」が獲得できる。しかし、機械的に「門司にフク井が存在する」ことを解析するには至らない。

そこで、本研究では、2文に分けて情報を述べたブログ文からパターンの対を半自動で獲得し、パターン辞書を構築する。パターン辞書を利用することで、機械的に2文から情報を解析できるように改良する。

# 第3章 パターン辞書の構築

本研究での、パターン辞書の構築方法を説明する。

## 3.1 パターン辞書構築の概要

ブログ文を入力とし、場所と存在物の抽出用パターン対を半自動で作成し、パターン辞書を構築する。パターン辞書を図 3.1 のフローで構築する。

## 3.2 構築方法

### 3.2.1 手順

図 3.1 における手順の流れを説明する。

手順 1：用意したブログ文から、連続する単文（動詞述語に限る）の対を自動で抽出する。

手順 2：抽出した単文対から、場所と存在物を含む単文対を、自動で選出する。

手順 3：選出した単文対に対して、存在物、場所および重要表現の注釈を人手で付与し、不適切な単文対を取り除く。こうして、タグ付きの単文対を得る。

手順 4：タグ付き単文対から、場所・存在物の分析用のパターン対を自動で作成する。

ただし、手順 2 では、係り受け解析ツール CaboCha[2] による出力で、LOCATION タグが付与されている名詞、または、CaboCha の意味属性に「地域」が付与されている名詞を、場所とする。日本語語彙大系で一般名詞意味属性が具体物である名詞を、存在物とする。また、パターン対に格を取り込むため、場所と存在物はそれぞれ、助詞の直前にある名詞のみを扱う。

CaboCha の出力例を図 3.2 に示す。「門司」という単語に LOCATION タグ、および、意味属性「地域」が付与されているため、「門司」を場所として扱う。

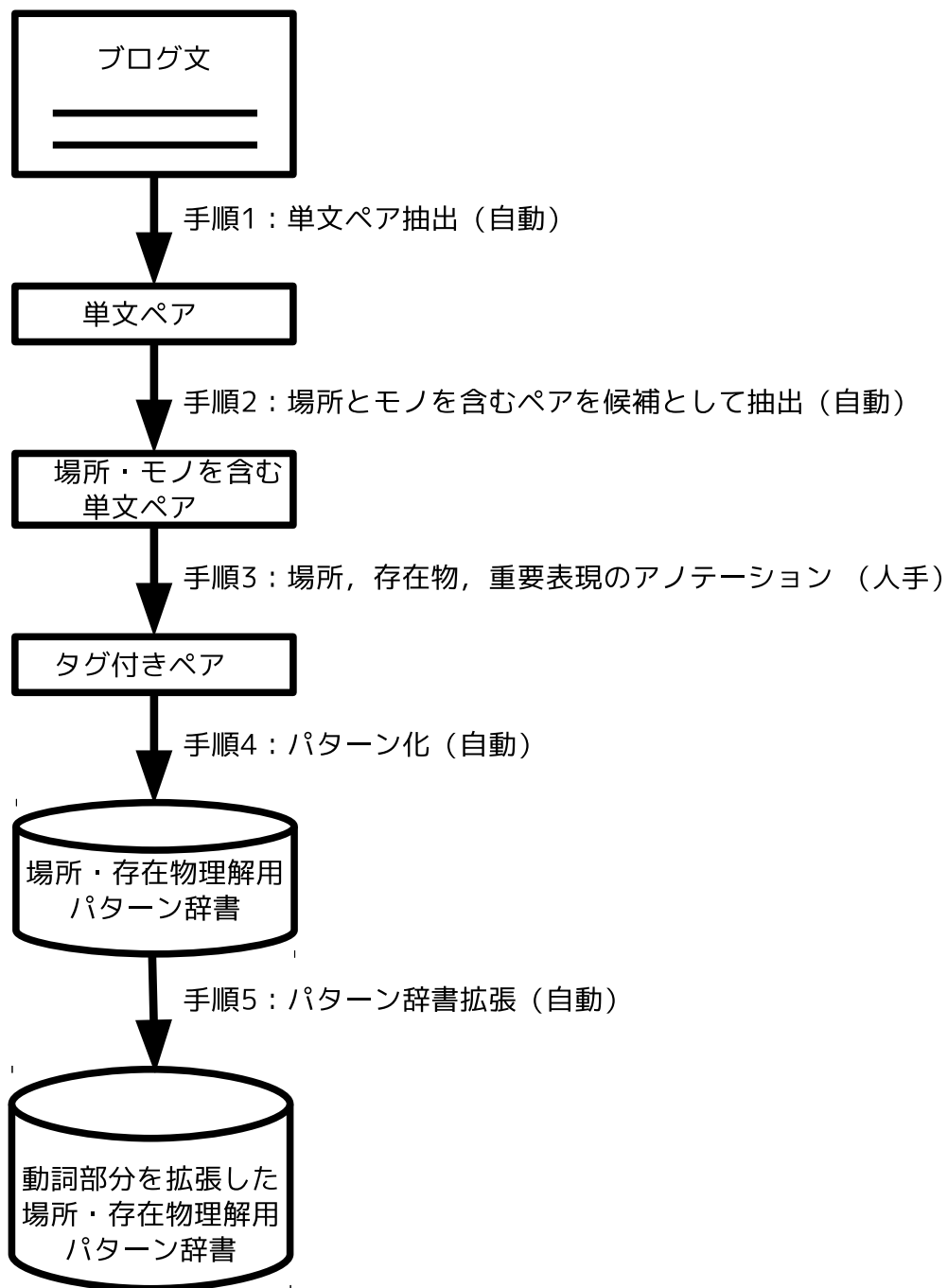


図 3.1 パターン辞書構築のフロー

---

門司へ行った。

0 1D 0/1 0.000000

門司 名詞, 固有名詞, 地域, 一般, \*\*, 門司, モジ, モジ B-LOCATION

へ 助詞, 格助詞, 一般, \*\*\*, へ, へ, エ O

1 -1D 0/1 0.000000

行っ 動詞, 自立, \*\*, 五段・カ行促音便, 連用タ接続, 行く, イッ, イッ O

た 助動詞, \*\*\*, 特殊・タ, 基本形, た, タ, タ O

。 記号, 句点, \*\*\*, \*, 。, 。, 。 O

EOS

フク井を食べた。

0 1D 1/2 0.000000

フク 名詞, 固有名詞, 人名, 名, \*\*, フク, フク, フク B-PERSON

井 名詞, 一般, \*\*\*, \*, 井, ドンブリ, ドンブリ O

を 助詞, 格助詞, 一般, \*\*\*, を, ヲ, ヲ O

1 -1D 0/1 0.000000

食べ 動詞, 自立, \*\*, 一段, 連用形, 食べる, タベ, タベ O

た 助動詞, \*\*\*, 特殊・タ, 基本形, た, タ, タ O

。 記号, 句点, \*\*\*, \*, 。, 。, 。 O

EOS

---

図 3.2 CaboCha の出力例

### 3.2.2 用いるタグセット

手順3において、タグ付けに用いるタグセットについて説明する。

- <N\_O 変数化部分 | 助詞等字面で残す部分 > : 存在物として扱う名詞を示すタグ
- <N\_L 変数化部分 | 助詞等字面で残す部分 > : 場所として扱う名詞を示すタグ
- <V lemmatize する部分 > : 動詞部分（重要表現）を示すタグ
- <N\_X 変数化部分 | 助詞等字面で残す部分 > : 文の表現構造の持つ意味を、パターン化後も保つため字面や変数にしたい部分（重要表現）を示すタグ

### 3.2.3 パターン辞書の拡張

半自動によるパターンの追加では、パターン辞書の規模が小さくなる。そのため、日本語語彙大系の用言意味属性を利用し、動詞部分を置き換えたパターンを追加することで、パターン辞書の拡張を行う。

## 3.3 パターン辞書の構築結果

パターン対の作成、および、パターン辞書の拡張により、パターン辞書の構築を行った。

### 3.3.1 パターン対の作成

3つのブログサイトから、約22ヶ月分のブログデータを取得し、パターン対の作成を行った。入力は、取得したブログ文約967,000,000文であった。作成したパターン対は59ペアとなった。

図3.1における各手順での入力の推移を以下に示す。

手順1：用意したブログ文約967,000,000文から、連続する単文の対約300,000件を、自動で抽出する。

手順2：抽出した単文対から、場所と存在物を含む単文対約900件を、自動で選出する。

手順3：選出した単文対に対して、存在物、場所および重要表現の注釈を人手で付与し、不適切な単文対を取り除く。こうして、タグ付きの単文対57件を得る。

手順4：タグ付き単文対から，場所・存在物の分析用のパターン対59ペアを自動で作成する．

手順3において，作成したタグ付き単文対の例を図3.3に示す．

- 
- (1) 週末に <N\_L:門司|まで> ドライブに <V:行っ> できました。  
<N\_O:フク井|を><V:食べ> できました。
- (2) 深川の <N\_L:富岡八幡|に> お参りに <V:行き> ました。  
私は半世紀この <N\_O:八幡様|に><N\_X:|お世話に><V:なっ> ています。
- 

図 3.3 タグ付き単文対の例

図 3.3 より作成したパターン対を図 3.4 に示す．

- 
- (1) FP:N1 まで/' 行く'[MDLC2]。  
意味属性制約:N1 (場所)  
SP:N1 を/' 食べる'[MDLC2]。  
意味属性制約:N1 (具体物)
- (2) FP:N1 に/' 行く'[MDLC2]。  
意味属性制約:N1 (場所)  
SP:N1 に/' お世話に/' なる'[MDLC2]。  
意味属性制約:N1 (具体物)
- 

図 3.4 パターン対の例

### 3.3.2 用言意味属性を利用したパターン辞書の拡張

3.2節の方法でパターン辞書を拡張した．FPは58パターンから約155,000パターンに，SPは59パターンから約138,000パターンに増加した．

図3.5に置き換えた動詞の例を示す．括弧内は用言意味属性を示す．

- 
- (1) 行く (18 物理的移動)  
集まる, 歩み寄る, 歩く, 急ぐ, 移動する, 他
- (2) 食べる (23 身体動作)  
会う, 仰ぐ, いただく, 外食する, 飲む, 他
- 

図 3.5 動詞置き換えの例

## 第4章 パターン辞書による場所と存在物の抽出

### 4.1 場所と存在物の組の抽出手順

図 4.1 に場所と存在物の組の抽出の流れを示す。手順を以下に示す。

手順 1：SPM にテスト文を入力し，形態素解析データを得る。

手順 2：形態素解析データを patlap に入力し，照合結果を得る。

手順 3：patlap から出力された照合結果より，場所と存在物の組を自動で抽出する。

### 4.2 パターン対による情報抽出

パターン対は 1 文目と照合するパターン (FP) と 2 文目と照合するパターン (SP) で構成する。入力を連続する単文とし，同時に，FP が入力文の 1 文目と，SP が入力文の 2 文目と適合した場合に，場所と存在物の組を抽出する。

2.2 節の，2 文に分けて情報を述べた例からは，図 4.2 に示すパターン対を用いることで場所と存在物の組を抽出できる。「/」は任意の文字列と「' 動詞'」は動詞の原形，および，活用形と，変数「[MDLC2]」は任意のモダリティ表現とそれぞれ適合する。

#### 4.2.1 パターン照合

パターン照合の方法について説明する。パターン照合には，patlap (本研究室で作成された情緒推定ツール) を用いる。パターン照合の入力は，入力文に連続する ID を付与し，SPM (形態素解析ツール) で形態素解析した出力とする。



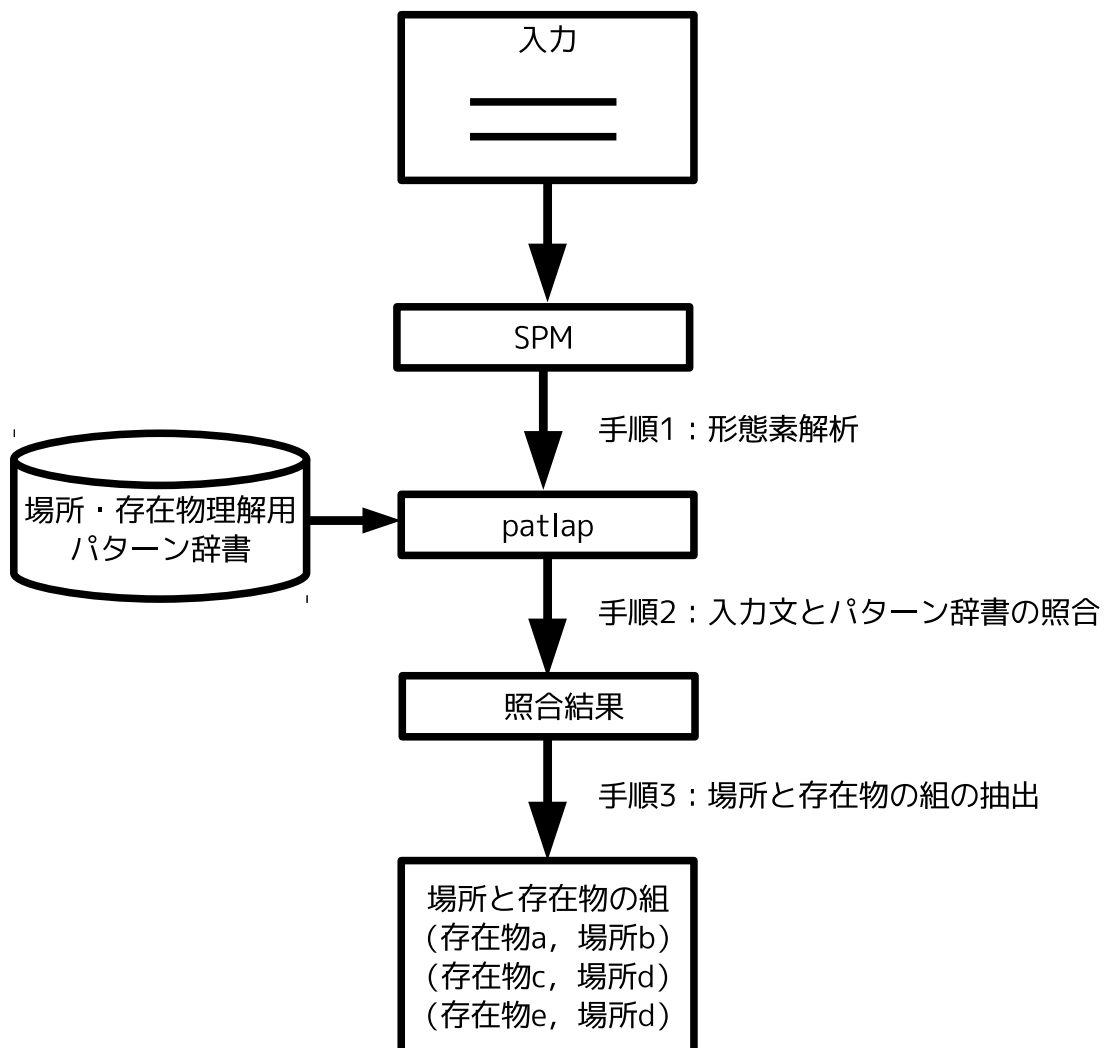


図 4.1 場所と存在物の組の抽出のフロー

---

(FP) 用言パターン：N1 へ/’ 行く’[MDLC2]。  
 意味属性制約：[N1 (場所)]

(SP) 用言パターン：N1 を/’ 食べる’[MDLC2]。  
 意味属性制約：[N1 (具体物)]

---

図 4.2 用言パターン対の例

## 4.2.2 patlap 用パターン対

パターン辞書に収録されるパターン対は，patlap で処理するために，以下の形式変換を要する．

- 1 文目パターン辞書と 2 文目パターン辞書に分割している．
- パターンペアを明示するため，パターン ID を付与している．
- 変数制約条件は，3.1.1 節の手順 2 の判定条件を用いて照合結果を絞り込むため，一律に  $N\#(*)$  とする．

以上の形式変換を行い，図 4.3 の形式をもつ，patlap 向けのパターンを作成する．ただし， $num$  はパターンによって抽出される語が場所であるか存在物であるかを表す数値である． $num = 1$  で場所， $num = 2$  で存在物を抽出するパターンであることを示す．FP に付与する SPID は，対になる SP の ID である．

---

FP：場所・存在物抽出パターン	<tab>	変数の制限	<tab>	num	<tab>
抽出される語	<tab>1<tab>	FPID	<tab>	SPID	(<tab>SPID…)
SP：場所・存在物抽出パターン	<tab>	変数の制限	<tab>	num	<tab>
抽出される語	<tab>2<tab>	SPID			

---

図 4.3 patlap 向けのパターン形式

図 3.4 の 2 組のパターン対を，patlap 向けのパターン形式に変換したパターンを図 4.4 に示す．

---

(1) FP: $N1$ まで /' 行く '[MDLC2]。	N1(*);MDLC2(*)	1	N1	1	1	1
SP: $N1$ を /' 食べる '[MDLC2]。	N1(*);MDLC2(*)	2	N1	2	1	
(2) FP: $N1$ に /' 行く '[MDLC2]。	N1(*);MDLC2(*)	1	N1	1	2	2
SP: $N1$ に /' お世話に /' なる '[MDLC2]。	N1(*);MDLC2(*)	2	N1	2	2	

---

図 4.4 パターン対の例

## 第5章 実装

本章では、パターン辞書を用いて場所と存在物の組を抽出するシステムの実装方法について説明する。

### 5.1 システムの概要

システムの実装を図 5.1 に示す。

#### 5.1.1 手順 1 から手順 5：パターン辞書の構築

手順 1 から手順 5 は、第 3 章のパターン辞書構築手順 (図 3.1)、および、パターン辞書の拡張と同一である。プログ文から、パターン対を作成し、辞書を構築する。その後、用言意味属性を用いて、パターン辞書を拡張する。

#### 5.1.2 手順 6：パターン辞書の形式変換

手順 6 では、4.2.2 節のパターンの形式変換を行い、パターン辞書を patlap 向けの形式に変換する。

#### 5.1.3 手順 7 から手順 9：2 文からの場所と存在物の組の抽出

手順 7 から手順 9 は、第 4 章のパターン照合、および、場所と存在物の組の抽出と同一である。patlap を用いて、形式変換したパターン辞書を読み込み、形態素解析を行った入力文との照合結果を得る。その後、照合結果から 2 文からの場所と存在物の組を抽出する。

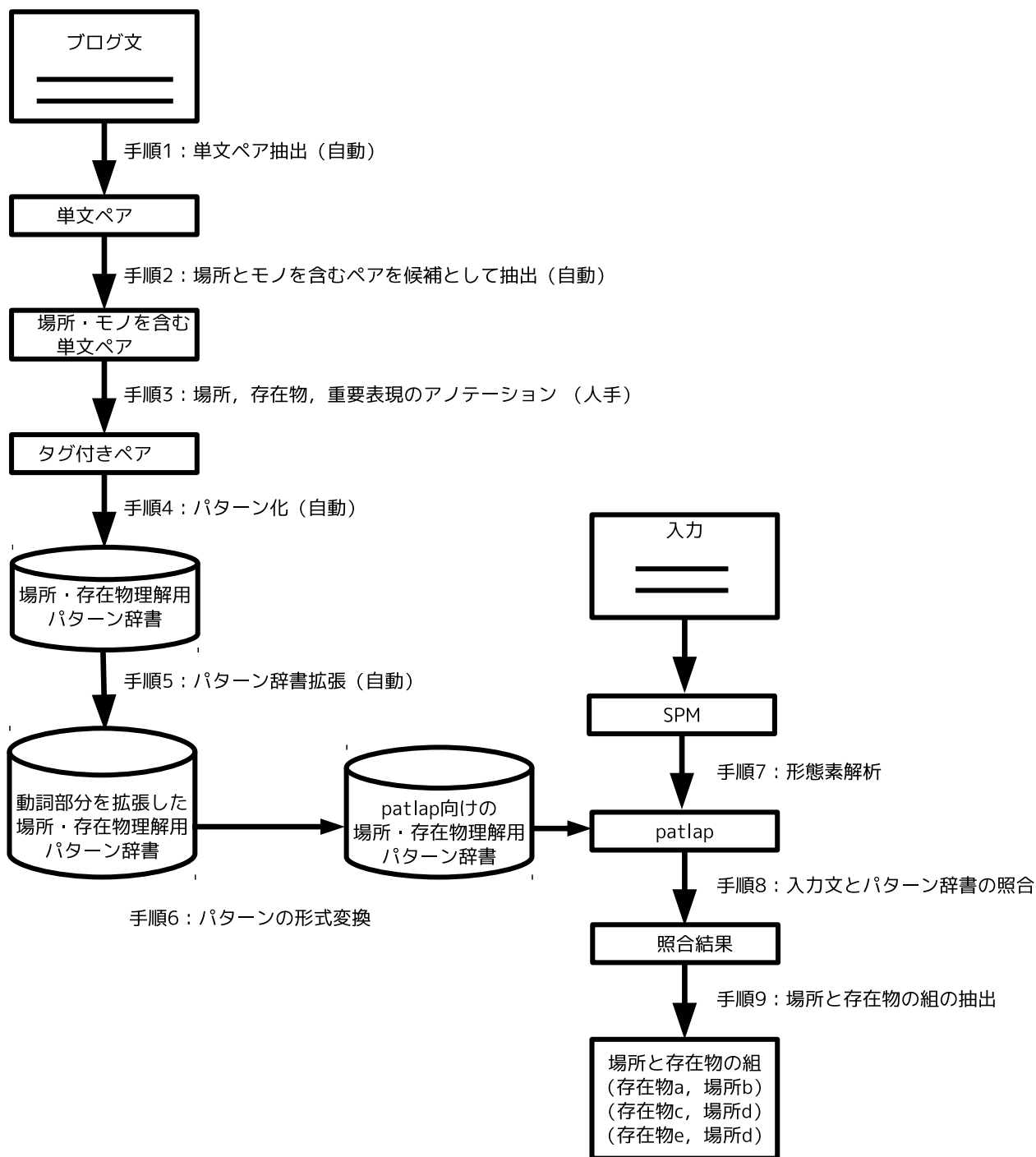


図 5.1 システムの実装

図 4.4 をパターン辞書として用い，入力を次に示す図 5.2 を形態素解析して得た図 5.3 とした際の patlap の出力を，図 5.4 に示す．図 5.4 を解析すると，入力文の ID とその文に対する出力から，2 文が図 4.4 の (1) のパターン対に適合していると判るため，場所と存在物の組（フク井，門司）を抽出する．

---

000001	週末に門司までドライブに行ってきました。
000002	フク井を食べてきました。

---

図 5.2 入力文の例

---

INPUT=000001=0xffff=0x7ff

週末に門司までドライブに行ってきました。

1. /週末 (1500, {NI:2681, NI:2708, KR:6503y27, KR:9001x45, IM:16920, IM:16970})
2. +に (7430)
3. /門司 (1900, {NI:5, NI:48, NK:67, NK:69})
4. +まで (7520)
5. /ドライブ (1210, {NI:1664, KR:2501a28, KR:5207h12, IM:15242})
6. +に (7430)
7. /行っ (2114, 行く, 行っ, {NY:18, NY:23, NY:15, NY:29, NY:32, NY:20, NY:5, KR:2700a00, KR:2700a01, KR:8908a02, IY:6970, IY:8410})  
(2394, 行なう, 行なっ, {NY:16, NY:29, NY:23, IY:2421, IY:3300, IY:5210})
8. +てき (2823, てくる, てき)
9. +まし (7234, ます, まし)
10. +た (7216)
11. +。 (0110)
12. /nil

=====

INPUT=000002=0xffff=0xff

フク井を食べてきました。

1. /フク (1920, {NI:49, NK:70})
2. +井 (1100, {NI:911, NI:905, NI:848, KR:0300p51, IM:13650, IM:13670})
3. +を (7430)
4. /食べ (2413, 食べる, 食べ, {NY:23, NY:2, KR:0003a20, KR:0300a00, IY:3210, IY:3233})
5. +てき (2823, てくる, てき)
6. +まし (7234, ます, まし)
7. +た (7216)
8. +。 (0110)
9. /nil

=====

---

図 5.3 SPM の出力の例

---

INPUT=000001

週末に門司までドライブに行ってきました。

[11, "。", "1", "門司", "1", "1", "1", nil, nil]

[11, "。", "1", "ドライブ", "1", "2", "2", nil, nil]

[11, "。", "1", "週末", "1", "2", "2", nil, nil]

INPUT=000002

フク井を食べてきました。

[8, "。", "2", "フク井", "2", "1", nil, nil, nil]

---

図 5.4 出力の例

# 第6章 2文からの場所と存在物の抽出 実験

3章で構築したパターン辞書を用いた，抽出実験について述べる．

## 6.1 実験の目的

場所と存在物の組を出力すべきか，出力しないべきかに注目し，実験を行う．パターン辞書を用いた場所と存在物の組の抽出結果を評価し，それを用いた場所と存在物の解析の有効性を示すことを目的とする．

## 6.2 テスト文

実験に使用するテスト文について説明する．3つのブログサイトから新たに8ヶ月分のブログデータを取得し，30ヶ月分のブログデータを用意した．1つ以上の場所と存在物の組を出力すべき文対26件，および，場所と存在物の組を出力すべきではない文対21件の，計47件の文対を用意した．この時，パターン作成に用いた文対とは異なる文対になるように選出した．

## 6.3 正解データ

テスト文には，人手で，文対毎に1つ以上の正解データを付与した．場所と存在物の組を出力すべき文対には，それぞれ，1つ以上の正解データを付与する．場所と存在物の組を出力しないべき文対には，それぞれ，正解データとして，空のペア ( $\phi$  のペア) を付与する．これにより，場所と存在物の組を出力すべき文対26件に，35組の正解データを付与した．また，場所と存在物の組を出力すべきではない文対21件に21件の正解データを付与した．正解集合の大きさは56となった．



## 6.4 実験方法

比較のため，次の3つの手法について実験を行う．

手法1：パターン辞書におけるパターン対が同時に2文対に適合する際に場所と存在物のペアを抽出

手法2：パターン対を分解し，FP と SP の全ての組み合わせで再構成したパターン辞書を用いて，場所と存在物のペアを抽出

手法3：パターン辞書を用いずに，3.1.1 節の手順2で用いた場所と存在物の判定条件で，場所と存在物のペアを抽出

手法1，手法2では図4.1の手順により，patlapを用いて，パターン辞書とテスト文の照合を行う．手法1では，patlapの照合結果から，パターン対のFPが文対の1文目と，SPが2文目と同時に適合していた場合，場所と存在物の組を抽出する．手法2では，それぞれ場所と存在物を抽出するFPとSPが，手法1と同様に文対と適合した場合，場所と存在物の組を抽出する．

各手法の出力は場所と存在物のペアであり，同様に場所と存在物の得られない場合の出力は $\phi$ ペアである．

## 6.5 評価方法

実験結果の評価は， $F$ 値で行い，各手法の $F$ 値を比較する．評価基準として広く用いられている適合率，再現率の値も算出する．

### 6.5.1 評価値

正解集合を $A$ とし，出力集合を $O$ とする．また，正解集合 $A$ と出力集合 $O$ の一致数を $|A \cap O|$ とする．それぞれの評価値は次のようになる．

適合率は，出力の正確性の指標である．

$$\langle \text{適合率} \rangle = \frac{|A \cap O|}{|O|} \quad (6.1)$$

再現率は，出力の網羅性の指標である．

$$\langle \text{再現率} \rangle = \frac{|A \cap O|}{|A|} \quad (6.2)$$

$F$  値は適合率と再現率の調和平均であり，総合性能の指標である．

$$\langle F \text{ 値} \rangle = \frac{2 \cdot \langle \text{適合率} \rangle \cdot \langle \text{再現率} \rangle}{\langle \text{適合率} \rangle + \langle \text{再現率} \rangle} = \frac{2 \cdot |A \cap O|}{|A| + |O|} \quad (6.3)$$

## 6.6 実験結果

6.3 節に示す正解データを付与した場合の結果を図 6.1 に示す．逆に，正解データとして  $\phi$  のペアを付与しなかった場合の結果を図 6.2 に示す．

表 6.1 6.3 節の正解データを場所と存在物の抽出結果

手法	出力数	正解数	一致数	適合率	再現率	$F$ 値
1	48	56	28	0.58	0.50	0.54
2	57	56	37	<b>0.65</b>	<b>0.66</b>	<b>0.65</b>
3	65	56	35	0.54	0.63	0.58

手法 2 で， $F$  値が最もよく，性能が良いといえる．図 6.2 と比べると，手法 3 の過剰な出力が結果に反映されている．

表 6.2  $\phi$  のペアを正解として付与しなかった場合の場所と存在物の抽出結果

手法	出力数	正解数	一致数	適合率	再現率	$F$ 値
1	10	35	8	<b>0.80</b>	0.23	0.36
2	29	35	20	0.69	0.57	0.63
3	65	35	35	0.55	<b>1.00</b>	<b>0.71</b>

手法 1 においては，出力数が少ないが，適合率が高い．間違った出力が考慮されていないため，手法 3 において，他手法と比べて  $F$  値が高い．

### 6.6.1 抽出例

テスト文 (図 6.1) と手法 1，手法 2 において適合したパターン (図 6.2)，および，出力された組と正解データ (表 6.3) を示す．この例では，それぞれの手法において，同一の出力が得られたため，まとめて示す．場所と存在物の組は「(存在物, 場所)」で示す．正しい情報を抽出しており，浅草ではおみくじを引けることが解析できる．

---

昨日，浅草に行ってきました。  
おみくじをひきました。

---

図 6.1 テスト文の例 3

---

(1) FP:N1 へ/’ 行く’[MDLC2]。  
意味属性制約:N1 (場所)  
SP:N1 を/’ 引く’[MDLC2]。  
意味属性制約:N1 (具体物)

---

図 6.2 手法 1，手法 2 において適合したパターン 1

### 6.6.2 誤り例 1

抽出された場所と存在物の組が，誤りであった際のテスト文（図 6.3）と手法 1，手法 2 において適合したパターン（図 6.4），および，出力された組と正解データ（表 6.4）を示す。この例では，それぞれの手法において，同一の出力が得られたため，まとめて示す。

意味属性制約により，高山が場所に適合しないため，図 6.3 の入力に対する正解は  $(\phi, \phi)$  となる。しかし，（バス，五箇山）という誤りの組が出力されている。図 6.4 のパターン対は，3.3.2 節の動詞の置き換えによる拡張で作成されたパターン対であるため，拡張による不利益が確認できる。

表 6.3 出力された場所と存在物の組 1

出力	正解
(おみくじ，浅草)	(おみくじ，浅草)

---

五箇山から白川郷経由で高山に戻ってきました。  
明日は奈良でシンポジウムなので、今からバスで名古屋まで向かいます。

---

図 6.3 テスト文の例 1

---

(1) FP:N1 から /' 戻る '[MDLC2]。  
意味属性制約:N1 (場所)  
SP:N1 で /' 向かう '[MDLC2]。  
意味属性制約:N1 (具体物)

---

図 6.4 手法 1, 手法 2 において適合したパターン 2

### 6.6.3 誤り例 2

手法 1 と比べて、手法 2 の出力で誤りの組が増えた際のテスト文 (図 6.5) と手法 1, 手法 2 において適合したパターン (図 6.6), および、出力された組 (表 6.5) を示す。

手法 1 では、図 6.6 の (1) のみが、手法 2 では、(1), (2), (3) が適合する。それぞれのパターン対は、3.3.2 節の動詞の置き換えによる拡張で作成されたパターン対であり、(2), (3) はパターン対の再構成により作成されたパターン対である。

手法 1 の出力から、動詞の置き換えによるパターン対の拡張の不利益、および、パターン辞書の規模の小ささに起因する網羅性の低さが確認できる。手法 2 の出力から、パターン対の再構成による不利益が確認できる。

表 6.4 出力された場所と存在物の組 2

出力	正解
(バス, 五箇山)	( $\phi$ , $\phi$ )

ずっと若い頃 友達と二人で 名古屋から寝台車で 広島まで行きました。  
早朝の平和公園で ベンチに座り 資料館が開くのを 2時間も待っていました。

図 6.5 テスト文の例 2

- (1) FP:N1 から /' 行く'[MDLC2]。  
意味属性制約:N1 (場所)  
SP:N1 で /' 待つ'[MDLC2]。  
意味属性制約:N1 (具体物)
- (2) FP:N1 から /' 行く'[MDLC2]。  
意味属性制約:N1 (場所)  
SP:N1 に /' 待つ'[MDLC2]。  
意味属性制約:N1 (具体物)
- (3) FP:N1 から /' 行く'[MDLC2]。  
意味属性制約:N1 (場所)  
SP:N1 が /' 待つ'[MDLC2]。  
意味属性制約:N1 (具体物)

図 6.6 手法 1, 手法 2 において適合したパターン 3

表 6.5 出力された場所と存在物の組 3

手法 1	手法 2	手法 3	正解
(平和公園, 名古屋)	(平和公園, 広島)	(平和公園, 広島)	(平和公園, 広島)
	(ベンチ, 広島)	(ベンチ, 広島)	(ベンチ, 広島)
	(資料館, 広島)	(資料館, 広島)	(資料館, 広島)
	(平和公園, 名古屋)	(平和公園, 名古屋)	
	(ベンチ, 名古屋)	(ベンチ, 名古屋)	
	(資料館, 名古屋)	(資料館, 名古屋)	

## 第7章 考察

### 7.1 手法1の性能が低い問題について

手法1では，表6.2や表6.5からわかるように，出力が得られない件数が多い．この原因は，テスト文に，パターン対の助詞部分が文対の1文目，および，2文目に適合する例が少なかったためである．このことから，パターン辞書の規模が小さく，拡充の必要性があるといえる．

### 7.2 動詞の置き換えによるパターン辞書の拡張について

6.6.2節，および，6.6.3節において，動詞の置き換えにより作成されたパターン対が適合し，誤りを含む場所と存在物の組の抽出を行った．動詞の置き換えに用いる用言意味属性の条件付け，辞書の拡張方法の再検討，および，作成・拡張したパターン対の見直しの必要性があるといえる．

### 7.3 手法2のパターン対の再構成について

6.6.3節の表6.5において，再構成によって誤り出力が増加することが確認できる．しかし，表6.1より，他手法より性能が良く，今回の様にパターン辞書が小さい場合，手法2のパターン対の拡張は有効であるといえる．

## 第8章 おわりに

本研究では、用言パターンを対にして用いることで、連続する2つの動詞述語文から、場所と存在物の組の抽出を行うためのパターン辞書の構築を行った。

場所と存在物の組が出力されるべきか、それとも、出力されないべきかについて注目し、実験を行った結果、テスト文の文対47件に対し、パターンを用いる手法の  $F$  値が0.54となり、比較手法の  $F$  値と比べて高くなり、場所と存在物の解析精度をわずかに向上させることを確認した。パターンを対にして用いる手法では、パターン辞書の規模が小さいため、満足な結果が得られず、有効性の確認に至らなかった。今後の課題は、パターン辞書の拡充によるパターン対の網羅性向上、および、パターン辞書の見直しによる抽出性能の正確性向上である。

# 謝辞

本研究を進めるにあたり，種々の御助言を頂きました村田真樹教授，および，村上仁一准教授に心から御礼申し上げます．

徳久雅人講師には，始終に渡り研究の進め方や本論文の書き方など，細部に渡るご指導を頂きました．ここに深く感謝いたします．

その他様々な場面でご助力をいただいた計算機工学 C 講座の皆様に感謝の意を表します．



## 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: “日本語彙大系”, 岩波書店, 1997.
- [2] CaboCha: Yet Another Japanese Dependency Structure Analyzer.  
<http://code.google.com/p/cabochoa/>