

概要

日本語語彙大系のパターンを用いたブログの解析が行われている。日本語語彙大系は用言述語文をカバーする。しかし、名詞述語文はカバーしていない。例えば、“ノドグロを食べる”という用言述語文からは情緒を推定することができるが、“ノドグロは境港産だ”という名詞述語文は無視されるため何の推定も行われぬ。そのため、ブログ文の意味解析に有用である文が無視されるため支障がある。今田 [1] は、生産物名詞を主語とする名詞述語文に着目し、主語と述語の意味関係の観点における意味構造の類型化を試み、「範疇叙述型」、「属性叙述型」、「外延叙述型」および「その他」の4つの類型に分類した。

そこで、本研究では、今田の分類をベースに名詞述語文の意味を解析するためのパターン辞書の作成を目的とする。そのために以下のことを行う。まず、生産物名詞が主語である名詞述語文の事例を基にタグの付与および型名の追記を行い、パターン辞書を作成する。タグの付与および型名の追記は人手で行う。パターンの作成を行った結果、29件のパターンが作成できた。次に、パターン照合による解析を行い、作成した辞書の性能を評価する。解析の際、入力文に対し、複数の適合結果が出力される場合があるので、適合結果の縮退および優先順位付けを行う。性能評価には、再現率、適合率および F 値を用いる。オープンテストの結果、再現率は 0.32、適合率は 0.05、 F 値は 0.09 と低い値となった。適合率が低い理由は、入力文に対する出力数が過剰であるためで、選択性能の強化が必要である。再現率が低い理由は、入力文に正しく適合するパターンがないためで、パターンの規模の拡大および、パターンの拡張が必要である。今後の課題として、パターン辞書の規模の拡大、パターンの拡張および適合結果の選択性能を強化が挙げられる。

目次

第1章	はじめに	1
第2章	名詞述語文	2
2.1	名詞述語文の意味分類	2
2.1.1	範疇叙述型	2
2.1.2	属性叙述型	3
2.1.3	外延叙述型	5
2.2	本研究のアプローチ	6
第3章	名詞述語文の文型パターン辞書の構築	7
3.1	手順	7
3.1.1	タグの付与および型名の追記	8
3.1.2	パターンの作成	9
3.1.3	パターンの統合	9
3.2	結果	10
第4章	名詞述語文の文型パターン辞書を用いた解析	20
4.1	手順	20
4.2	入力文にパターンを照合	21
4.3	縮退および優先順位付け	22
4.3.1	縮退	22
4.3.2	上位語・下位語の関係	22
4.4	適合結果の選択について	23
4.5	解析の様子	24
第5章	実験	27
5.1	実験の目的	27

5.2	実験方法	27
5.2.1	評価規準	27
5.2.2	使用辞書	28
5.3	パターンの適合実験	29
5.3.1	準備	29
5.3.2	結果	29
5.4	型名の判定実験	30
5.4.1	準備	30
5.4.2	結果	30
5.5	意味解析の実験	31
5.5.1	準備	31
5.5.2	結果	31
第6章	考察	33
6.1	実験結果に関する考察	33
6.2	辞書に関する考察	33
6.3	誤り解析	34
6.3.1	形態素解析の誤り	34
6.3.2	型の定義・3つ組の不足	35
6.3.3	パターンの形不足	35
6.4	上位語・下位語の関係をを用いた優先付けの問題点	36
第7章	おわりに	37

目 次

3.1	パターンの構築の流れ	7
4.1	パターン辞書を用いた解析の流れ	20

表 目 次

2.1	EAV モデルを用いた属性名詞述語文の意味構造	4
2.2	EAV モデルに基づく構文類型	4
3.1	タグセット	8
3.2	品詞の一部	8
3.3	関数の一部	8
5.1	辞書 1 のパターン適合の性能	29
5.2	辞書 2 のパターン適合の性能	29
5.3	辞書 3 のパターン適合の性能	29
5.4	辞書 1 の型名判定の性能	30
5.5	辞書 2 の型名判定の性能	30
5.6	辞書 3 の型名判定の性能	30
5.7	辞書 1 の意味解析の性能	31
5.8	辞書 2 の意味解析の性能	31
5.9	辞書 3 の意味解析の性能	31

第1章 はじめに

日本語語彙大系のパターンを用いたブログの解析が行われている [2][3] . 日本語語彙大系は用言述語文をカバーする [4] . しかし, 名詞述語文はカバーしていない . 例えば, “ノドグロを食べる” という用言述語文からは情緒を推定することができるが, “ノドグロは境港産だ” という名詞述語文は無視されるため何の推定も行われぬ . そのため, ブログ文の意味解析に有用である文が無視されるため支障がある .

今田 [1] は, 生産物名詞を主語とする名詞述語文に着目し, 主語と述語の意味関係の観点における意味構造の類型化を試み, 「範疇叙述型」, 「属性叙述型」, 「外延叙述型」および「その他」の 4 つの類型に分類した . そこで, 本研究では, 今田の分類をベースに名詞述語文の意味を解析するためのパターン辞書の作成を目的とする .

本論文の構成は以下の通りである . 第 2 章では, 名詞述語文について述べる . 第 3 章では, パターン構築の流れを述べる . 第 4 章では, パターン辞書を用いた解析の流れを述べる . 第 5 章では, パターン照合によるパターン辞書の性能評価の実験を行う . 第 6 章では, 本研究の考察を行う . 最後に第 7 章でまとめを述べる .

第2章 名詞述語文

本章では、まず名詞述語文が伝える意味について今田 [1] の分類を引用して述べる。次に本研究のアプローチを述べる。

2.1 名詞述語文の意味分類

名詞述語文は、述語の品詞が名詞である文のことをいう。その中で生産物名詞が主語である名詞述語文は、その他を含む4通りに分類される [1]。

2.1.1 範疇叙述型

名詞述語文の主要な類型の1つは、主語名詞句が表す事物の帰属する範疇を述語名詞句が述べるものである。今田は以下の例文を示した。

- (1) トルコの「汗と絨毯」は、トルコ東南部の町、ウルファで昔ながらの方法で絨毯を作る職人たちを追った ドキュメンタリー だ。
- (2) お正月の獅子舞は、我々を守ってくれる神の「権現」であり、村境に張った「しめ縄」は、悪霊の侵入を防ぐ 道具 だった。
- (3) ワゴン車 はレンタル会社から借り出された もの で、借用書類からの指紋採取が進んでいた。
- (4) ところが、同委員長は「石油 はわが国の主権の シンボル である」として、真っ向から対立。

(1, 2) は述語名詞句が範疇的概念を表す名詞述語文の例であるが, (1) のように作品といった類の下位種を表す語が現れる場合や, (2) のように比較的上位の類概念を表す語が現れる場合, さらには (3) のようにごく抽象的なレベルの範疇しか表さない形式名詞が現れる場合などがある。

範疇的概念の中には, 事物の博物学的分類を表すものもあれば, そうではないものもある。例えば (1) の「ドキュメンタリー」は博物学的分類の一種であるが, (4) の「シンボル」は「石油」の博物学的分類を示しているわけではない。(4) の述語も範疇的概念を表すものではあるが, 事物の博物学的分類を表しているというよりは, 何らかの目的や機能を共有する事物や概念のグループを表している。

範疇を述べる文には, 範疇情報自体が文の主要な伝達内容であるものもあるが, 範疇情報自体よりも, それを修飾する部分の方が文の主要な伝達内容であるものも多い。(1) では, 作品や商品がどのようなカテゴリーに分類されるかということが問題となっており, 「ドキュメンタリー」といった範疇情報自体が文の主要な伝達内容となっている。一方, (2, 3) では, 「道具」および「もの」は文の主要な伝達内容ではなく, 「どのような」にあたる部分の情報が文の主要な伝達情報となる。「どのような」にあたる部分の情報には様々な種類のものがある。例を挙げると, (2) の「悪霊の侵入を防ぐ」は, 事物の目的を表す。

以上に対し, 本研究では次のようにエッセンスを取り出す。範疇叙述型は, 主語が述語に対して下位語, 述語が範疇(上位語), および, 追加の情報を示す語句として追加情報が解析で得られる。

範疇叙述型の例

原文: パソコンはただの道具だ。

下位語: パソコン

上位語: 道具

追加情報: ただの

2.1.2 属性叙述型

範疇叙述型では述語名詞の主要部が範疇的概念を表す語である場合を見たが, 次のような事例では述語は主語が表す事物の帰属する範疇を表しているわけではなく, 味, 色, 図柄, 材質, 数量, 価格などの属性的概念を表している。今田は以下の例文を示した。

(5) 偽造投票用紙 は本物と同じ薄い 緑色 だが, 紙がやや薄めで朱色で印刷された文字や公印が不鮮明。

(6) 日本の海運会社が運航している 外航貨物船 はおよそ 二千隻 である。

名詞述語文によって叙述される属性の種類は多様であり，それらをどのように整理するかは課題である．

構文的特徴についてもう少し詳しく区別すると，属性を表す述語名詞句の主要部は，属性の種類を表す場合と，値を表す場合がある．コンピュータ・データベースなどで用いられる EAV モデルというデータモデルで記述すると，これらの文の意味構造は表 2.1 のように記述することができ，属性または値のいずれかが述語名詞句の主要部になっているということになる．

実体，属性，値の三項の現れ方には，いくつかのパターンがある．代表的なものとしては，表 2.2 のような構文が考えられる．E+A/V 型の構文においては，主語は実体ではなく属性にシフトしており，文全体は後述する外延叙述型の構文（いわゆる役割 - 値文）にシフトしている．

一般的に言えば，範疇叙述型の文が主語（対象）と述語（範疇）の二項関係であるのに対して，属性叙述型の文は EAV モデルで構造化されるような三項関係で記述することができる．

以上に対し，本研究では次のようにエッセンスを取り出す．属性叙述型は，主語が実体を表し，述語がその属性となる．また，追加の情報を示す語句は値となるが，述語が属性および値となる場合もある．解析で実体，属性，値が得られる．

属性叙述型の例

原文:このカレーは 300 円 だ。

実体:カレー

属性:300 円（価格）

値:300 円

表 2.1: EAV モデルを用いた属性名詞述語文の意味構造

実体 (Entity)	属性 (Attribute)	値 (Value)
偽造投票用紙	色	緑色
外航貨物船	数量	二千隻

表 2.2: EAV モデルに基づく構文類型

構文	例
E/V+A	[_E 東京タワー] は [_V 333m] の [_A 高さ] だ。
E+A/V	[_E 東京タワー] の [_V 高さ] は [_A 333m] だ。
E/A/V	[_E 東京タワー] は [_V 高さ] が [_A 333m] だ。
E/A	[_E 東京タワー] は [_A 333m] だ。

2.1.3 外延叙述型

名詞述語文のある種のもは、主語に内包的概念を表す語を取って、述語でその外延を叙述する。今田は以下の例文を示した。

- (7) 宇宙研の主力 ロケット が今回の M3S2 で、宇宙開発事業団は「きく6号」を打ち上げたH2ロケットだ。
- (8) 掲げる 旗 は 民主主義 だ。
- (9) そう、二階建てで洋風の 建物 は広小路の名古屋郵便局と ここだけ でした。

引用

これは従来の研究では倒置指定文や役割 - 値文といった名称で知られる文類型に相当する。役割を表す名詞には、字義通りの意味で用いられるものもあればやや比喩的な意味で用いられるものもある。

外延を述べる文の中には、特定少数の要素を値として持つような役割概念ではなく、不特定多数の要素を外延として持ち得るような範疇的概念が主語の位置に現れる事例(9)もある。従って外延叙述型の文は、必ずしも主語が役割を表すとは限らないのであるが、どのような場合にこのような構文が可能になるかについては現段階では十分に明らかではない。

以上に対し、本研究では次のようにエッセンスを取り出す。外延叙述型は、内包的概念を表す語を主語が上位語、述語が外延(下位語)、および、追加の情報を示す語句として追加情報が解析で得られる。ただし、述語が外延を述べるため、追加情報が得られない場合がある。

外延叙述型の例

原文:使った道具はドライバーだ。

上位語:道具

下位語:ドライバー

追加情報:使った

2.2 本研究のアプローチ

今田 [1] の研究から，生産物を主語とする名詞述語文として「範疇叙述型」，「属性叙述型」および「外延叙述型」の分類があることがわかった．この3通りの分類を基に，本研究では「範疇叙述型」の文に対して下位語，上位語および追加情報の3つ組の情報を，「属性叙述型」の文に対して実体，属性および値の3つ組の情報を，そして「外延叙述型」の文に対して上位語，下位語および追加情報の3つ組の情報をそれぞれ抽出することが本質的と考えた．表 2.2 にある通り，構文類型に着目した意味解析に可能性が示された．これを受けて，3つ組情報の抽出可能なパターンをあらかじめ試作しておき，生産物を主語とする名詞述語文の型の分類および3つ組情報の抽出による解析を機械化することを試みる．

第3章 名詞述語文の文型パターン辞書の構築

名詞述語文の分類および情報の抽出を行うパターン辞書を，名詞述語文の事例から生成する．

3.1 手順

まず，名詞述語文に，変数，関数，字面および品詞を表すタグを付与し，型名を追記する．次に，タグより変数化，関数化，字面化を行いパターンを作り，タグおよび型名に基づき3つ組を追加する．3つ組は，意味解析結果として出力する情報である．最後に，包含関係にある複数のパターンを1つのパターンに統合する．図3.1にパターンの構築手順を示す．

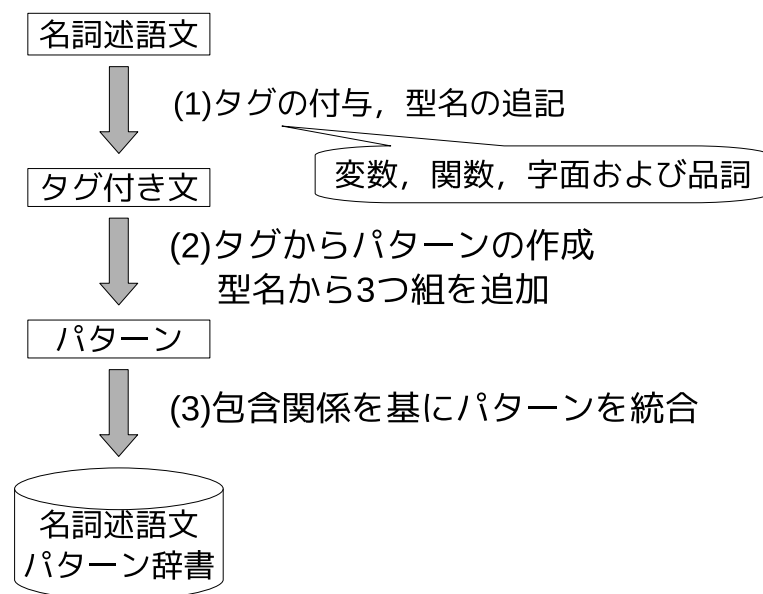


図 3.1: パターンの構築の流れ

次節には各処理の詳細を示す．

3.1.1 タグの付与および型名の追記

名詞述語文は，ID 番号が付けられている．ID 番号はアルファベット + 6 桁の番号とする．文にスペースで分かち書きしておく．文にタグを付与し，型名を追記する．タグは変数，関数，字面および品詞を表す．処理後は次の形式とする．

タグセットは「 $\langle \rangle$ 」内に，タグセット名，品詞，関数，字面およびパターンに残す字面を表記する．タグセット名は表 3.1 のアルファベットである．品詞は表 3.2 を表記する．関数（表 3.3）は不要である場合は表記する必要はない．関数および字面の間は「:」で区切る．字面は変数に変換する単語もしくは句である．パターンに残す字面がある場合は，その前に「|」を表記しておくことで実現している．

表 3.1: タグセット

タグ	意味
$\langle E : \dots \rangle$	実体
$\langle A : \dots \rangle$	属性
$\langle V : \dots \rangle$	値
$\langle U : \dots \rangle$	上位語
$\langle L : \dots \rangle$	下位語
$\langle I : \dots \rangle$	追加情報

表 3.2: 品詞の一部

品詞	意味
<i>N</i>	名詞
<i>V</i>	動詞
<i>NUM</i>	数詞
<i>AJ</i>	形容詞

「*P*」が末尾にある場合「句」となる．

表 3.3: 関数の一部

関数	意味
<i>.dantei</i>	判定詞
<i>.kako</i>	過去
<i>.reru</i>	受動態・可能態
<i>.#da</i>	だ，である，です
<i>^rentai</i>	連体修飾

—— タグ付けの例 ——

a000001 皮が薄い分、たっぷり詰まった〈E N:つぶしあん | は〉〈V V.kako:抑えた〉〈A N.dantei:甘さで〉、さっぱりした仕上がり。属性叙述型

本研究ではタグ付けの処理を人手で行った。

3.1.2 パターンの作成

タグを付与した文からパターンを作成する。タグより変数化，関数化および字面化を行いパターンを作り，タグおよび型名に基づき3つ組を追加する。パターンの変数の前には記号「/」を追加する。記号「/」は任意の文節と適合する離散記号である。また，パターン，型名および3つ組の他に，各変数の制約条件およびID番号を追加する。ただし，本研究で制約条件は設定していない。

—— パターンの作成の例 ——

タグ付き文:

a000001 皮が薄い分、たっぷり詰まった〈E N:つぶしあん | は〉〈V V.kako:抑えた〉〈A N.dantei:甘さで〉、さっぱりした仕上がり。属性叙述型

パターン:

/N1 は/V2.kako/N3.dantei N1(*);V2(*);N3(*) id:a000001 属性叙述型 entity:N1 attribute:N3 value:V2

3.1.3 パターンの統合

包含関係にある複数のパターンを1つのパターンに統合する。型名の一致しているパターン同士が包含関係にあった場合，包含関係を考慮した上でパターンの統合を行う。統合する際「;」区切りで統合前のパターンのID番号を追記する。本研究ではパターンの統合の処理を人手で行った。

—— パターンを統合した例 ——

統合前:

/N1 は/VP2.kako/N3.dantei id:a000024 範疇叙述型

/N1 は/VP2.kako/N3.dantei.kako id:a000025 範疇叙述型

/N1 は/VP2.kako/N3.dantei.kako id:a000029 範疇叙述型

統合後:

/N1 は/VP2.kako/N3.dantei[.kako] id:a000024;id:a000025;id:a000029 範疇叙述型

例は、制約条件および3つ組情報を省略している。統合前の3つのパターンは、パターン末尾に「.kako」の有無だけの違いのため包含関係にある。統合後は「.kako」を任意化を行う。任意化は「[]」で関数などを囲むことのできる。また、ID番号を3つ表記している。

統合後、変数の前にある任意の文節と適合する記号「/」を、副詞・形容詞の文節に適合するよう記号「/」に変更する。その際、パターンが原文と適合しないように変更をしてはならない。また、末尾の判定詞「.dantei」もしくは過去「.kako」の任意化を行う。

変更例

変更前：

/N1は/VP2.kako/N3.dantei[.kako] id:a000024;id:a000025;id:a000029 範疇叙述型

変更後：

/N1は/fkVP2.kako/fkN3.dantei[.kako] id:a000024;id:a000025;id:a000029 範疇叙述型

3.2 結果

文献 [1] で示された例文 (41 文) から、29 件のパターンを収録した辞書を作成した。以下に、辞書のレコードを示す：

原文：皮が薄い分、たっぷり詰まったつぶしあんは抑えた甘さで、さっぱりした仕上がり。

レコード

パターン：/N1は/fkVP2/fkN3[.dantei]

ID 番号：a000001

型名：属性叙述型

3つ組：実体:N1, 属性:N3, 値:VP2

原文 1：偽造投票用紙は本物と同じ薄い緑色だが、紙がやや薄めで朱色で印刷された文字や公印が不鮮明。

原文 2：また、宴会客のたばこの火の不始末の疑いも浮上したが、その後の調べでステージはたばこの火などでは燃えにくい材質であることも分かった。

レコード

パターン：/N1 は /fkAJP2^rentai/fkN3[.dantei]

ID 番号：a000002;a000004

型名：属性叙述型

3 つ組：実体:N1, 属性:N3, 値:AJP2

原文：衣装一つをとっても黒縹子に松竹梅の模様に肩から袖にかけて金糸よりのしめ飾り、その下が火焰太鼓と桜の絵柄、前に垂れた帯は鯉の滝昇りの図である。

レコード

パターン：/N1 は /fkNP2/fkN3[.dantei]

ID 番号：a000003

型名：属性叙述型

3 つ組：実体:N1, 属性:N3, 値:NP2

原文 1：タンクはグラスファイバー製で上部に直径五十センチのネジ込み式のふたがある。

原文 2：実験器具は実費で、参加費は二万五千 三万円。

レコード

パターン：/N1 は /fkN2[.dantei]

ID 番号：a000005;a000010

型名：属性叙述型

3 つ組：実体:N1, 属性:N2, 値:N2

原文 1：日本の海運会社が運航している外航貨物船はおよそ二千隻である。

原文 2：日本電子機械工業会が 11 日に発表した昨年 11 月のカラーテレビ・VTR 統計によると、国内出荷台数は、カラーテレビが前年同月比 7・8% 増の 81 万 8000 台、VTR は 12・0% 増の 51 万 7000 台で、ともに 4 カ月連続して前年実績を上回った。

原文 3：カレーは三百円だが、サラダは三百十円する。

レコード

パターン：/N1 は /fkNUM2[.dantei]

ID 番号：a000006;a000007;a000009

型名：属性叙述型

3 つ組：実体:N1, 属性:NUM2, 値:NUM2

原文：建設省のキャッチフレーズになった「住宅」は、九五年度予算、九六年度要求予算ともに、一七・六%と同じシェアだ。

レコード

パターン：/N1 は [,] /fkVP2 /fkN3[.dantei]

ID 番号：a000008

型名：属性叙述型

3 つ組：実体:N1, 属性:N3, 値:VP2

原文：「免許行政下で一物一価だったビールが、一物多価の時代に突入した」節目の年だった。

レコード

パターン：/NP1.dantei.kako /fkN2 が [,]

ID 番号：a000011

型名：属性叙述型

3 つ組：実体:N2, 属性:NP1, 値:NP1

原文 1：トルコの「汗と絨毯」は、トルコ東南部の町、ウルファで昔ながらの方法で絨毯を作る職人たちを追ったドキュメンタリーだ。

原文 2：同副首相は七日夜のロシア・テレビとのインタビューで、現在ドゥダエフ政権部隊が使用している大量の兵器は、一九九一年にロシア軍がチェチェンを撤退した後に同政権に引き渡したものだと言明。

原文 3：バッジは、実行グループの一人が落としたものであることも裏付けられている。

レコード

パターン：/N1 は [,]/fkVP2.kako/fkN3[.dantei]

ID 番号：a000012;a000026;a000028

型名：範疇叙述型

3 つ組：下位語:N1，上位語:N3，追加情報:VP2

原文：発泡酒はアルコール飲料だが、麦芽使用率が六五%以下のため、日本の酒税法の分類ではビールと見なされず、通常のビールより低い税率が適用される。

レコード

パターン：/N1 は/fkNP2 だが、/fkNP3 の

ID 番号：a000013

型名：範疇叙述型

3 つ組：下位語:N1，上位語:NP2，追加情報:NP3

原文 1：お正月の獅子舞は、我々を守ってくれる神の「権現」であり、村境に張った「しめ縄」は、悪霊の侵入を防ぐ道具だった。

原文 2：阪神間を結ぶ道路は、生活用品や復旧資材を運ぶ貴重な幹線だからだ。

レコード

パターン：/N1 は [,]/fkVP2 の/fkN3[.dantei][.kako]

ID 番号：a000014;a000020

型名：範疇叙述型

3 つ組：下位語:N1，上位語:N3，追加情報:VP2

原文：ここ一、二年のうちに朝の食卓に登場してきたオレンジジュースも新しい市場を開拓した商品である。

レコード

パターン：/N1も/fkVP2/fkN3[.dantei]

ID 番号：a000015

型名：範疇叙述型

3つ組：下位語:N1，上位語:N3，追加情報:VP2

原文1：ワゴン車はレンタル会社から借り出されたもので、借用書類からの指紋採取が進んでいた。

原文2：写真で首に巻いているマフラーは、インタビューの直前に歌手の荻野目洋子さんにプレゼントされたものである。

レコード

パターン：/N1は[、]/fkVP2.reru.kako/fkN3[.dantei]

ID 番号：a000016;a000023

型名：範疇叙述型

3つ組：下位語:N1，上位語:N3，追加情報:VP2

原文：核の傘は日米安保体制の根幹にかかわる問題だ。

レコード

パターン：/N1は[、]/fkVP2^rentai/fkN3[.dantei]

ID 番号：a000017

型名：範疇叙述型

3つ組：下位語:N1，上位語:N3，追加情報:VP2

原文：ラップは若者へのメッセージだ。

レコード

パターン：/fkN1は/fkN2への/fkN3[.dantei]

ID 番号：a000018

型名：範疇叙述型

3つ組：下位語:N1，上位語:N3，追加情報:N2

原文：その半面、バブルの温床だった商業地は、大都市圏で二ケタの大幅マイナスとなった。

レコード

パターン：/fkN1 の/fkN2.dantei.kako/fkN3 は

ID 番号：a000019

型名：範疇叙述型

3 つ組：下位語:N3，上位語:N2，追加情報:N1

原文：ところが、同委員長は「石油はわが国の主権のシンボルである」として、真っ向から対立。

レコード

パターン：/N1 は/fkNP2 の/fkN3[.dantei]

ID 番号：a000021

型名：範疇叙述型

3 つ組：下位語:N1，上位語:N3，追加情報:NP2

原文 1：これについて、冷蔵庫の査察を拒否した工場は、その冷蔵庫にある製品は国内向けのものだったからだと弁明している。

原文 2：公邸は大使のものではない。国民の財産です。

レコード

パターン：/N1 は/fkN2 の/fkN3[.dantei][.kako]

ID 番号：a000022;a000030

型名：範疇叙述型

3 つ組：下位語:N1，上位語:N3，追加情報:N2

原文1：アウシュビッツで公開されているガス室は解放後に共産党政権がでっち上げたもので、ナチス指導者はユダヤ人絶滅を計画したことなどなかった、と指摘している。

原文2：男は陸軍士官学校卒の現役中尉で、銃は同士官学校から盗み出したものだった。

原文3：凶器はドライバーの柄に千枚通しのようなものをつけたものだった。

レコード

パターン：/N1は/fkVP2.kako/fkN3[dantei][.kako]

ID番号：a000024;a000025;a000029

型名：範疇叙述型

3つ組：下位語:N1，上位語:N3，追加情報:VP2

原文：一カ月にわたるロシア軍の進攻に激しく抵抗しているドゥダエフ・チェチェン政権の兵器が旧ソ連軍の残したもので、ドゥダエフ大統領との兵器引き渡し交渉に、グラチョフ・ロシア国防相があたっていたことがロシアで問題化している。

レコード

パターン：/N1が/fkVP2.kako/fkN3[dantei]

ID番号：a000027

型名：範疇叙述型

3つ組：下位語:N1，上位語:N3，追加情報:VP2

原文：太閤秀吉の城であった伏見城は、家康の持ち城となっていたからである。

レコード

パターン：/fkN1の/fkN2.dantei.kako/fkN3は[、]

ID番号：a000031

型名：範疇叙述型

3つ組：下位語:N3，上位語:N2，追加情報:N1

原文：製薬メーカー側は「分からない」「うちの製品ではない」と回答を避けるケースもあるが、抗がん剤が自社の製品だった場合「分からない」では不自然。

レコード

パターン：/fkN1 が/fkN2 の/fkN3[.dantei][.kako]

ID 番号：a000032

型名：範疇叙述型

3 つ組：下位語:N1，上位語:N3，追加情報:N2

原文：宇宙研の主力ロケットが今回のM3S2で、宇宙開発事業団は「きく6号」を打ち上げたH2ロケットだ。

レコード

パターン：/N1 が/fkN2 の/fkN3[.dantei]

ID 番号：a000033

型名：外延叙述型

3 つ組：上位語:N1，下位語:N3，追加情報:N2

原文：ともに使われた凶器は、けん銃だった。

レコード

パターン：/N1 は [,]/fkN2[.dantei][.kako]

ID 番号：a000034

型名：外延叙述型

3 つ組：上位語:N1，下位語:N2，追加情報:NULL

原文1：商品は一年物スーパー定期で、10万円につき一回の抽選権がある。

原文2：掲げる旗は民主主義だ。

レコード

パターン：/N1 は/fkN2[.dantei]

ID 番号：a000035;a000038

型名：外延叙述型

3 つ組：上位語:N1，下位語:N2，追加情報:NULL

原文：コークの造るアメリカキューブの秘密兵器は何なのか。

レコード

パターン：/NP1のN2は/fkN3[.dantei]

ID番号：a000036

型名：外延叙述型

3つ組：上位語:N2，下位語:N3，追加情報:NP1

原文：第二の落とし穴は、私たちの神経はエネルギーをブドウ糖に頼っており、B1の欠乏が神経に重大な障害をもたらすことです。

レコード

パターン：/N1は[、]/VP2/fkN3.#da

ID番号：a000037

型名：外延叙述型

3つ組：上位語:N1，下位語:N3，追加情報:VP2

原文：新方式の引き金は「学校名や序列を安上がりの選抜の手段として利用し、有力大学出身者以外を寄せ付けない大手企業の採用システムは、受験競争の過熱に重大な一役を買ってきた」との中教審中間報告だったという。

レコード

パターン：/N1は/VP2との/fkN3[.dantei][.kako]

ID番号：a000039

型名：外延叙述型

3つ組：上位語:N1，下位語:N3，追加情報:N2

原文：陽性の食物は地面の下で育つ野菜、海草、大豆、小魚などで、陰性の食物は夏に取れる野菜や果物、ビールなど。

レコード

パターン：/N1は/N2

ID番号：a000040

型名：外延叙述型

3つ組：上位語:N1，下位語:N2，追加情報:NULL

原文：そう、二階建てで洋風の建物は広小路の名古屋郵便局とここだけでした。

レコード

パターン：/N1は/NP2と/N3

ID番号：a000041

型名：外延叙述型

3つ組：上位語:N1，下位語:N3，追加情報:N2

第4章 名詞述語文の文型パターン辞書を用いた解析

4.1 手順

まず、パターン辞書の各パターンを入力文に照合する。適合結果として型名や3組情報が得られる。次に、適合した複数の結果の縮退および順位付けを行う。1つの文に複数のパターンが適合する場合、上位語・下位語の関係を満たす適合結果を優先する。さらに、3組における主名詞の組および型名が同一の適合結果は1つの結果としてまとめる。最後に、優先順位の高い適合結果を選択し、入力文の意味解析結果として型名および3組情報を出力する。図4.1にパターン辞書を用いた解析手順を示す。

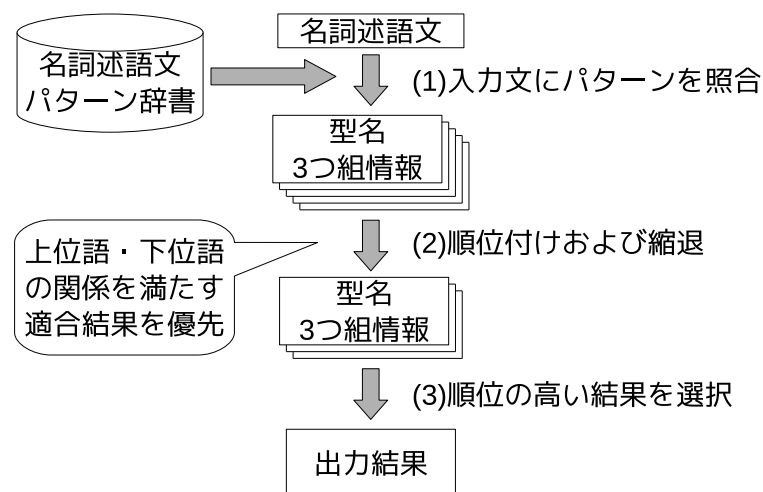


図 4.1: パターン辞書を用いた解析の流れ

次節には各処理の詳細を示す。

4.2 入力文にパターンを照合

パターン照合を行う。ツールはSPM（形態素解析）およびPatlap（本研究室で作成された情緒推定ツール）を用いる。SPMでの形態素解析結果を基にPatlapでパターン照合を行う。以下に形態素解析結果およびPatlapでのパターン照合結果の例を示す。

形態素解析結果

```
INPUT=a000006=0x3ffffff=0xffff
日本の海運会社が運航している外航貨物船はおよそ二千隻である。
1. /日本 (1940,{NI:463,NI:385,NK:30})
2. +の (7410)
3. /海運 (1100,{NI:1982,NI:1980,KR:6200h15,IM:15289})
4. +会社 (1100,{NI:374,NI:428,KR:4308h19,IM:11420,IM:12120})
5. +が (7410)
6. /運航し (2233, 運航する, 運航し,{NY:18,KR:2700b02,IY:8610})
7. +ている (2817)
8. /外航 (1210,{NI:2669,IM:168D0})
9. +貨物船 (1100,{NI:989,KR:6200k41,IM:13690})
10.+は (7530)
11./およそ (6170)
12.+二千 (1610, 2千, 2千)
13.+隻 (6220,{NI:2595,NI:2588,KR:1602z64})
14.+で (7255, だ, で)
15.+ある (2786)
16.+。 (0110)
17./nil
```

パターン照合結果の例

```
INPUT=a000006
日本の海運会社が運航している外航貨物船はおよそ二千隻である。
[13, "隻", "id:a000006;id:a000007;id:a000009", "属性叙述型", "entity:9", "attribute:13", "value:13"]
```

4.3 縮退および優先順位付け

4.3.1 縮退

適合した複数の結果の縮退を行う。まず、適合結果の3つ組情報のそれぞれの単語または句の最後の単語を残し、それ以外の単語を取り除く。

3つ組情報の処理の例

処理前:

```
[11, "もの", "id:a000016;id:a000023", "範疇叙述型", "lower:1,2", "upper:11", "additional_info:4,5,6,7,8"]
```

処理後:

```
[11, "もの", "id:a000016;id:a000023", "範疇叙述型", "lower:2", "upper:11", "additional_info:8"]
```

例の項目は、左から順にパターンに適合した最後の単語番号、その字面、適合したパターンのID番号、型名、3つ組情報（下位語、上位語、追加情報）である。処理後の3つ組情報は、それぞれ最後の単語番号のみ出力されるようにした。最後に、各3つ組情報および型名が一致した適合結果を1つの結果としてまとめる。

4.3.2 上位語・下位語の関係

日本語語彙大系の一般意味属性大系の木構造を用いて、主語および述語の上位語・下位語の関係から「範疇叙述型」もしくは「外延叙述型」かどうかの判定を行う。形態素解析結果を利用し、適合したパターンの主語が述語と比較して下位語であれば「範疇叙述型」、上位語であれば「外延叙述型」とし、パターンに付与されている型名と一致している場合は、一致した適合結果を優先する。

上位語・下位語の関係の例

バッジは、実行グループの一人が落としたものであることも裏付けられている。

上記の範疇叙述型の例では、主語「バッジ」、述語「もの」である。「バッジ」および「もの」の上位語・下位語の関係は、「バッジ」が「もの」に対して下位語で、「もの」は上位語となる。範疇叙述型は主語が下位語、述語は上位語という関係なため、この例文には正しく上位語・下位語の関係があると言える。

4.4 適合結果の選択について

上位語・下位語の関係を利用した優先順位付けを行ったとしても適合結果が複数ある場合、本研究では適合結果の1つを選択せず複数の結果を出力する。

適合結果の選択の例

入力文：

男は陸軍士官学校卒の現役中尉で、銃は同士官学校から盗み出したものだった。

適合結果：

1：[21, "もの", "id:a000024;id:a000025;id:a000029", "範疇叙述型", "lower:1", "upper:21", "additional_info:19"]

2：[21, "もの", "id:a000024;id:a000025;id:a000029", "範疇叙述型", "lower:12", "upper:21", "additional_info:19"]

上記の例では、入力文1文に対し、適合結果は2つ出力された。解析を行う際は、正解データと比較して型名および3つ組それぞれが一致しているか調べる。すべて一致していれば解析は成功と言える。

4.5 解析の様子

「ワゴン車はレンタル会社から借り出されたもので、借用書類からの指紋採取が進んでいた。」を入力文とした解析の流れを示す。

step1：入力文の形態素解析

形態素解析結果

INPUT=a000016=0x1ff=0xfffff

ワゴン車はレンタル会社から借り出されたもので、
借用書類からの指紋採取が進んでいた。

1. /ワゴン (1100,{NI:898,NI:988,KR:6313k56,IM:13670,IM:13690})
2. +車 (6280,{NI:988,NI:954,KR:4910k05,KR:6313k01,KR:6313k13,
KR:3200m11})
3. +は (7530)
4. /レンタル (1100,{NI:1910,KR:4213a23,IM:1527C})
5. +会社 (1100,{NI:374,NI:428,KR:4308h19,IM:11420,IM:12120})
6. +から (7410)
7. /借り (2413, 借りる, 借り,{NY:19,NY:29,KR:4214a00,IY:5730,IY:8720})
8. +出さ (2732, 出す, 出さ)
9. +れ (7113, れる, れ)
- 10.+た (7217)
- 11./もの (1100,{NI:706,NI:5,NI:2491,NI:1064})(1800)
- 12.+で (7410)(7255, だ, で)
- 13.+、 (0210)
- 14./借用 (1220,{NI:1911,KR:4214a01,KR:5400a79,IM:1527C})
- 15.+書類 (1100,{NI:1113,NI:919,KR:2200m05,IM:13670,IM:14400})
- 16.+から (7410)
- 17.+の (7410)
- 18./指紋 (1100,{NI:2584,NI:627,KR:9601h33,IM:13250,IM:16670})
- 19.+採取 (1220,{NI:1972,KR:4607a17,IM:15289})
- 20.+が (7410)
- 21./進む (2174, 進む, 進む,{NY:20,NY:18,NY:23,KR:4902a00,IY:2421,IY:5620,
IY:6970,IY:7410,IY:8300,IY:8310})
- 22.+でい (2813, でいる, でい)
- 23.+た (7216)
- 24.+。 (0110)
- 25./nil

step2 : 入力文とパターン辞書の照合

「ワゴン車はレンタル会社から借り出されたもので、借用書類からの指紋採取が進んでいた。」

(正解：範疇叙述型 (ワゴン車, もの, レンタル会社から借り出さ))

適合したパターンの一部

```
/N1 は [, ]/fkVP2.reru.kako/fkN3[.dantei] N1(*);VP2(*);N3(*)  
id:a000016;id:a000023 範疇叙述型 lower:N1 upper:N3 additional_info:VP2  
/N1 は/N2 N1(*);N2(*) id:a000040 外延叙述型 upper:N1 lower:N2 addi-  
tional_info:NULL
```

適合結果の一部

```
[11, "もの", "id:a000016;id:a000023", "範疇叙述型", "lower:1,2", "upper:11",  
"additional_info:4,5,6,7,8"]  
[11, "もの", "id:a000040", "外延叙述型", "upper:1,2", "lower:11", "addi-  
tional_info:NULL"]
```

step3 : 縮退および優先順位付けを行う

縮退後の結果の一部

```
[11, "もの", "id:a000016;id:a000023", "範疇叙述型", "lower:2", "upper:11",  
"additional_info:8"]  
[11, "もの", "id:a000040", "外延叙述型", "upper:2", "lower:11", "addi-  
tional_info:NULL"]
```

日本語語彙大系において「車」は下位語、「もの」は上位語なので範疇叙述型では関係を満たす。

優先された結果

```
[11, "もの", "id:a000016;id:a000023", "範疇叙述型", "lower:2", "upper:11",  
"additional_info:8"]
```

step4 : 正解データと照合

優先された結果の型名は「範疇叙述型」で、3つ組は、単語番号より左から「車」、「もの」、「出さ」である。正解データは、型名が「範疇叙述型」で、3つ組は、「ワ

ゴン車」、「もの」、「レンタル会社から借り出さ」であり、3つ組のそれぞれ最後の単語を見た場合に一致している。そのため、入力文に対して解析は成功している。

第5章 実験

5.1 実験の目的

本研究で作成したパターン辞書の解析性能の評価を行うことを目的とする。

5.2 実験方法

クローズドテスト(41文, 正解データ41件), および, オープンテスト(50文, 正解データ50件)をそれぞれ行う。オープンテストは, google 検索でキーワード「観光ブログ」から得たブログ文のうち, 生産物名詞を主語とする名詞述語文とする。

5.2.1 評価規準

一致数は, 正解データの型名および3つ組のそれぞれと, 適合結果に出力される正解データの型名および3つ組のそれぞれが一致した場合, 正解とし, 一致数にカウントする。再現率 R , 適合率 P , および, F 値を用いて, 本辞書を用いた3つ組および型名の解析性能を評価する。

再現率 R , 適合率 P , および, F 値は以下の式を用いる。

$$\text{再現率 } R = \frac{\text{一致数}}{\text{正解データ数}} \quad (5.1)$$

$$\text{適合率 } P = \frac{\text{一致数}}{\text{出力数}} \quad (5.2)$$

$$F \text{ 値} = \frac{2RP}{R+P} \quad (5.3)$$

5.2.2 使用辞書

パターンを入力文に照合すると、適合結果が複数出力される場合がある。適合結果が多い場合は辞書の性能が低下する。そのためパターンにある記号 / および /*fk* の変化が辞書にどのような影響を与えるか調査する。辞書は以下の3つの辞書を使用する。

辞書 1: /*fk* を使用せず変数の前に / を表記してある辞書

辞書 2: 辞書 1 の / を可能な限り /*fk* に変換を行った辞書

辞書 3: 可能な限り / および /*fk* を削除した辞書

辞書 2, 3 は辞書 1 から人手で変更した。変更した例を次に示す。

変更した辞書の例

辞書 1:

/N1 は /NP2 の /N3[.dantei] N1(*);NP2(*);N3(*) id:a000003 属性叙述型 entity:N1
attribute:N3 value:NP2

辞書 2:

/N1 は /*fk*NP2 の /*fk*N3[.dantei] N1(*);NP2(*);N3(*) id:a000003 属性叙述型
entity:N1 attribute:N3 value:NP2

辞書 3:

/N1 は NP2 の N3[.dantei] N1(*);NP2(*);N3(*) id:a000003 属性叙述型 entity:N1
attribute:N3 value:NP2

5.3 パターンの適合実験

入力文に対し，パターンが適合するかどうか実験を行う．

5.3.1 準備

辞書 1, 2, 3 を使用する．この実験ではパターンが適合しているかどうかの実験のため，入力文に対する出力はすべて 1 で，パターンが適合したかどうかを判定する．

5.3.2 結果

すべての辞書のクローズドテストにおける一致数は 41 件となった．オープンテストでは，辞書 1, 2 が一致数 37 件で F 値は 0.74，辞書 3 は一致数 36 件で F 値は 0.72 だった．以上を表 5.1, 5.2, 5.3 に示す．

表 5.1: 辞書 1 のパターン適合の性能

テスト名	R	P	F 値	一致	出力
クローズドテスト	1.00	1.00	1.00	41	41
オープンテスト	0.74	0.74	0.74	37	50

表 5.2: 辞書 2 のパターン適合の性能

テスト名	R	P	F 値	一致	出力
クローズドテスト	1.00	1.00	1.00	41	41
オープンテスト	0.74	0.74	0.74	37	50

表 5.3: 辞書 3 のパターン適合の性能

テスト名	R	P	F 値	一致	出力
クローズドテスト	1.00	1.00	1.00	41	41
オープンテスト	0.72	0.72	0.72	36	50

5.4 型名の判定実験

入力文に対し，型名が一致しているか実験を行う．

5.4.1 準備

辞書 1, 2, 3 を使用する．適合結果の 3 つ組情報は使用しないため無視する．型名は 3 通り使用しているので入力文に対して，最大で 3 つの出力がある．

5.4.2 結果

すべての辞書のクローズドテストにおける一致数は 39 件となった．再現率 R は 0.95(39/41)，適合率 P は 0.51(39/78) で F 値は 0.66 だった．オープンテストでは，一致数 37 件で一致数はすべて 32 件で，再現率 R は 0.64(32/50) となった．辞書 1 の出力数は 102 件で，適合率 P は 0.31(32/102) で F 値は 0.42 だった．辞書 2 の出力数は辞書 1 より 6 件減少し，適合率 P は 0.34(32/96) で F 値は 0.44 だった．辞書 3 の出力数はさらに 1 件減少し，適合率 P は 0.34(32/95) で F 値は 0.44 となった．以上を表 5.4, 5.5, 5.6 に示す．

表 5.4: 辞書 1 の型名判定の性能

テスト名	R	P	F 値	一致	出力
クローズドテスト	0.95	0.51	0.66	39	78
オープンテスト	0.64	0.31	0.42	32	102

表 5.5: 辞書 2 の型名判定の性能

テスト名	R	P	F 値	一致	出力
クローズドテスト	0.95	0.51	0.66	39	78
オープンテスト	0.64	0.34	0.44	32	96

表 5.6: 辞書 3 の型名判定の性能

テスト名	R	P	F 値	一致	出力
クローズドテスト	0.95	0.51	0.66	39	78
オープンテスト	0.64	0.34	0.44	32	95

5.5 意味解析の実験

入力文に対し，型名および3つ組情報が一致するか実験を行う．

5.5.1 準備

辞書1, 2, 3を使用する．パターン適合し，かつ型名および3つ組情報が正解データと一致したか判定する．

5.5.2 結果

すべての辞書のクローズドテストにおける一致数は37件であり，再現率 R は0.90(37/41)である．また，オープンテストにおける一致数は16件であり，再現率 R は0.32(16/50)である．

クローズドテストにおける辞書1の出力数は594件であり，適合率 P は0.06(37/594)， F 値は0.11である．辞書2では出力数が251件減少した．よって，適合率 P は0.10(37/343)， F 値は0.18である．辞書3では出力数がさらに20件減少し，適合率 P は0.11(37/323)， F 値は0.20である．一方，オープンテストにおける辞書1の出力数は536件であり，適合率 P は0.03(16/536)， F 値は0.05である．辞書2では出力数が243件減少した．よって，適合率 P は0.10(16/293)， F 値は0.09である．辞書3では出力数がさらに10件減少し，適合率 P は0.11(16/283)， F 値は0.10である．以上を表5.7, 5.8, 5.9に示す．

表 5.7: 辞書1の意味解析の性能

テスト名	R	P	F 値	一致	出力
クローズドテスト	0.90	0.06	0.11	37	594
オープンテスト	0.32	0.03	0.05	16	536

表 5.8: 辞書2の意味解析の性能

テスト名	R	P	F 値	一致	出力
クローズドテスト	0.90	0.10	0.18	37	343
オープンテスト	0.32	0.05	0.09	16	293

表 5.9: 辞書3の意味解析の性能

テスト名	R	P	F 値	一致	出力
クローズドテスト	0.90	0.11	0.20	37	323
オープンテスト	0.32	0.06	0.10	16	283

上位語・下位語の関係をを用いた優先付けを行ったが、オープンテストにおいて主語および述語の上位語・下位語の関係を満たすものはなかった。

第6章 考察

6.1 実験結果に関する考察

表 5.7 ,5.8 ,5.9 クローズドテストおよびオープンテストそれぞれの適合率 P は低い値となった。理由としては、クローズドテストおよびオープンテストの入力文に対する出力数が約 10 倍と過剰に出力したためである。そのため今後、適合結果の選択性能を強化する必要がある。

また、オープンテストの再現率 R が低い値であった。この原因は、オープンテストの入力文に適合するパターン、主にパターンの形もしくは 3 つ組の不足が見られた。そのためパターンの規模の拡大および既存のパターンの拡張が必要である。

6.2 辞書に関する考察

表 5.7 ,5.8 の出力数は、使用した辞書を辞書 1 から辞書 2 にした場合、クローズドテストでは 251 件、オープンテストで 243 件減少した。一致数に両辞書では変化は見られなかった。よって、辞書としては変数前の文節は任意のものでなく、副詞・形容詞の文節に限定した方が辞書の性能として高いと言える。

同様に、表 5.8 ,5.9 では、出力数は使用した辞書を辞書 2 から辞書 3 にした場合、クローズドテストでは 20 件、オープンテストで 10 件減少した。一致数に変化はなかった。辞書 3 は辞書 2 と比較するとほとんど変化はなかった。ただし、辞書 2 は辞書 3 と比較すると、入力文に対してパターンが適合する範囲が広い。よって、辞書としては性能に変化が見られなかったため辞書 2 の方がより適していると言える。

6.3 誤り解析

6.3.1 形態素解析の誤り

形態素解析の誤り例

入力文：公邸は大使のものではない。国民の財産です。
適合パターン：/N1は/N2 外延叙述型
得られた3つ組：公邸，大使，NULL

上記の例は，範疇叙述型として述語に「もの」が適合して欲しい文である．しかし，形態素解析結果を見ると，「もの」ではなく「ものではない」と区切られている．「もの」の品詞は名詞だが，「ものではない」の品詞は名詞ではない．そのため形態素解析結果の誤りにより正しくパターンが適合しなかった．

形態素解析結果

```
INPUT=a000030=0x1ffff=0x7ff
公邸は大使のものではない。国民の財産です。
借用書類からの指紋採取が進んでいた。
1. /公邸 (1100,{NI:447,NI:865,KR:2408k33,IM:12130,IM:13660})
2. +は (7530)
3. /大使 (1100,{NI:266,NI:323,KR:4306s36,KR:6208s28,IM:11260,IM:11270})
4. +の (7410)
5. /ものではない (3106)
6. +。 (0110)
7. /国民 (1100,{NI:154,KR:6805s02,IM:11242})
8. +の (7410)
9. /財産 (1100,{NI:1189,NI:1190,KR:7908k00,IM:14A50})
10.+です (7246)
11.+。 (0110)
12./nil
```

6.3.2 型の定義・3つ組の不足

型の定義・3つ組の不足の例

入力文：チヂミはモチモチで、イカの入った魚介チヂミでした。
適合パターン 1:/N1は/fkVP2.kako/fkN3[.dantei][.kako] 範疇叙述型
得られた3つ組 1：チヂミ，魚介チヂミ，モチモチで、イカの入っ
適合パターン 2:/N1は/N2 外延叙述型
得られた3つ組 2：チヂミ，魚介チヂミ，NULL

上記の例は外延叙述型として主語である「チヂミ」，述語に「魚介チヂミ」，追加情報として「モチモチで、イカの入っ」とパターンが適合して欲しい文である。しかし，適合パターン 1 は型名が違っているため，3つ組情報も違い，見かけ上3つ組が正しくても情報としては間違っている。適合パターン 2 は型名は正しい。だが，3つ組情報に欠落がある。このような3つ組の不足もしくは型の定義不足が原因の誤りが考えられる。そのため適合パターン 1 の形で型名の違うレコードがあればこの入力文は正しく適合する。

6.3.3 パターンの形不足

パターンの形不足の例

入力文：カウンターが4席ほどと、テーブル席が2席。
適合パターン：なし

上記の例は属性叙述型として主語である「カウンター」，述語に数量を表す「4席」とパターンが適合して欲しい文である。しかし，使用した辞書のパターンでは適合することはなかった。理由としてこの文は主語の後に格助詞「が」が使用されている。作成したパターンに格助詞「が」が使われているものは29件中4件だった。そのうち例のように主語，述語のみ適合するパターンはなかった。このようなパターンの形不足による誤りが考えられる。そのため助詞「は」および格助詞「が」のそれぞれどちらでも適合するような拡張が必要だと考えられる。

6.4 上位語・下位語の関係をを用いた優先付けの問題点

上位語・下位語の関係を満たすパターンを優先したが、オープンテストでは条件を満たすものがなかった。名詞述語文はそもそも新たに上位語・下位語の関係を定義する文として用いられる。

—— 上位語・下位語の関係を満たさない例 ——

製薬メーカー側は「分からない」「うちの製品ではない」と回答を避けるケースもあるが、抗がん剤が自社の製品だった場合「分からない」では不自然。

範疇叙述型の例を挙げると、主語の「抗がん剤」が下位語、述語である「製品」が上位語の関係となる。しかし、日本語語彙大系では「抗がん剤」および「製品」に上位語・下位語の関係を満たさない。これは、この例文で新たに「抗がん剤」および「製品」に、上位語・下位語の関係を定義したためであり、既知の情報ではない。そのため、この優先付けの効果が乏しいと考える。

第7章 おわりに

本研究では、名詞述語文の意味解析用のパターン辞書の作成を目的とした。今田 [1] は、名詞述語文の中でも生産物名詞が主語である名詞述語文に対し「範疇叙述型」「属性叙述型」および「外延叙述型」と分類した。本研究では、3つの分類を基に、生産物名詞が主語である名詞述語文の3つ組情報を抽出可能なパターンとして、29件を試作し、パターン照合で型および3つ組情報の出力による解析を試みた。変数の前を副詞・形容詞の文節に限る場合の辞書の性能は、クローズドテストの再現率は0.90、適合率は0.10、 F 値は0.18だった。オープンテストでの再現率は0.32、適合率は0.05、 F 値は0.09だった。

適合率が低い理由は、入力文に対する出力数が過剰であるためで、選択性能の強化が必要である。オープンテストの再現率が低い理由は、入力文に正しく適合するパターンがないためで、パターンの規模の拡大および、パターンの拡張が必要である。また、上位語・下位語の関係をを用いた優先順位付けを行ったが、名詞述語文はそもそも新たに上位語・下位語の関係を定義する文として用いられるので、効果は乏しかった。

今後の課題として、パターン辞書の規模の拡大、パターンの拡張および適合結果の選択性能を強化が挙げられる。

謝辞

本研究を進めるに当たり，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．

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

その他様々な場面で御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] 今田水穂:“日本語名詞述語文の類型と主語の意味分類について:京都大学テキストコーパスと分類語彙表を用いた調査・検査”, 筑波大学文藝・言語学系, 文藝言語研究, 言語篇, 60 巻, pp.25-48, 2011.
- [2] 徳久雅人, 奥村秀人, 村田真樹:“観光開発支援のためのブログ記事からの評判分析, 観光と情報”, 第7巻, 第1号, pp.85-98, 2011.
- [3] Masato Tokuhisa, Jin'ichi Murakami, Satoru Ikehara:“Affective Blog Analyzer - What People feel to”, Proceedings of the 2nd International Conference on Agents and Artificial Intelligence(ICAART2010), Vol.1, pp.247-252, 2010.
- [4] 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:“日本語語彙大系”, 岩波書店, 1997.