

概要

本研究では、パターン辞書を用いた名詞述語文の意味解析の方法について議論する。この方法では、入力文をパターンと照合すると、複数の照合結果が得られる。その中から照合結果の選択を行うことで、意味が解析されたことになる。パターン辞書について、用言述語文を対象とするものとして日本語語彙大系が既に存在する。重文複文を対象とするパターン辞書として鳥バンクが存在する。名詞述語文を対象とするパターン辞書は存在しない。既存のパターン辞書では網羅されていない文型が存在するので、その文型を網羅するパターン辞書を作成する必要がある。次に、パターン運用する場合、照合結果の選択が必要である。単文の場合は、主語および述語の意味属性コードを利用するが、重文複文では、主部や述部の具体性の判定において意味属性コードだけでは不足しており、主語の前の連体修飾表現の有無などの文構造の確認を加えて行う必要があるため、照合結果の選択条件が複雑になる。そこで、本研究では、名詞述語文の意味分類を確認し、名詞述語文を対象とするパターン辞書を作成するとともに、照合結果の選択条件を判定するために意味属性コードを利用する方法、および、機械学習を利用する方法を試みる。

以上について具体的に、まず、名詞述語文の意味分類を確認のため、名詞述語文の先行研究における例文、および、Wikipedia の文を対象に文の構造について分析を行う。次に、分析結果を基にパターン化基準を作成し、パターン辞書のエントリを作成する。作成した結果、例文 41 文から 21 エントリを作成できた。次に、パターンの照合結果の選択を行う方法を開発する。選択は、「意味属性コードを利用した照合結果の選択」の方法、および、「機械学習を利用した照合結果の選択」の方法の 2 つを提案する。

評価実験の結果、クローズドテストおよびクロスバリデーションでは、再現率、適合率、および、 F 値の全てが、「意味属性コードを利用した照合結果の選択」が上回り、オープンテストでは、再現率は「意味属性コードを利用した照合結果の選択」が、適合率は「機械学習を利用した照合結果の選択」がそれぞれわずかに上回った。誤りの主な原因は、単語を固有名詞とみなす処理が不足していたことだった。固有表現解析の問題として解決が可能と考えると、本研究により名詞述語文の意味解析のためのパターン辞書の構築および運用について方法を示すことができたといえる。

以上により，本研究では，名詞述語文の分類を，範疇叙述型，属性叙述型，および，外延叙述型に分類とすることに問題ないことを確認した．パターン化の基準を示したので，パターン辞書の構築が可能になった．名詞述語文の例文を基に 41 文から 21 エントリを得た．パターンの運用は，実験により，パターンの網羅性は良好であり，照合結果の選択は，現状では意味属性コードを利用する方法，および，機械学習を利用する方法のどちらも適応できることを確認した．

目次

第1章	はじめに	1
1.1	背景	1
1.2	関連研究	2
1.2.1	日本語語彙大系	2
1.2.2	鳥バンク	2
1.2.3	名詞述語文の分類	2
1.3	目的	3
第2章	今田による名詞述語文の分類	4
2.1	範疇叙述型	4
2.2	属性叙述型	5
2.3	外延叙述型	7
2.4	その他	8
第3章	名詞述語文の分析	9
3.1	目的	9
3.2	項目	9
3.3	今田の3分類の是非	9
3.4	主語についての抽出範囲	10
3.5	述語についての抽出範囲	11
3.6	「はが構文」の扱い	11
3.7	定形表現について	12
第4章	パターンの作成	13
4.1	パターン化基準	13
4.2	作成の様子	15
4.2.1	パターンの作成	15

4.2.2	エントリの作成	17
4.3	作成結果	19
4.3.1	範疇叙述型	20
4.3.2	属性叙述型	30
4.3.3	外延叙述型	36
第5章	パターンの運用	41
5.1	運用手順	41
5.2	実装	41
5.3	運用種類	41
5.4	意味属性を利用した照合結果の選択	41
5.4.1	原文との意味的な近さ	42
5.4.2	実行の様子	43
5.5	機械学習を利用した照合結果の選択	47
5.5.1	使用ツール	47
5.5.2	実行の様子	47
第6章	実験	50
6.1	実験の目的	50
6.2	実験の条件	50
6.3	クローズドテスト	51
6.4	クロスバリデーション	52
6.5	オープンテスト	52
第7章	考察	53
7.1	2つの手法の比較	53
7.2	作成したパターン辞書	53
7.3	「意味属性を利用」の方法のクローズドテスト	54
7.4	誤り分析	54
7.4.1	意味解析の成功例	54
7.4.2	固有名詞の処理不足	56
7.5	今後の対応すべき課題	57

目 次

3.1 「はが構文」と「非はが構文」の例	11
5.1 真理値ベクトルのイメージ	47

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 ([6] より引用)	6
2.2	EAV モデルに基づく構文類型 ([6] より引用)	7
3.1	Wikipedia 頻度上位単語	12
4.1	パターン化基準	14
4.2	エントリの内訳	19
5.1	素性生成ルール	48
6.1	型の内訳 (今田の例文)	50
6.2	型の内訳 (「自動車」の文)	50
6.3	型の内訳 (「潜水艦」の文)	51
6.4	今田の例文および「自動車」文の意味解析の性能 (クローズド)	51
6.5	今田の例文および「自動車」文の意味解析の性能 (クロスバリデーション)	52
6.6	「潜水艦」文の意味解析の性能 (オープン)	52

第1章 はじめに

1.1 背景

自然言語処理において、パターンを利用した処理が行われている。ここで言うパターンは、言語表現の構造を表すパターンであり、文字や変数で記述される。パターンは、文中に存在する意味を取り出すことができると考えられている。意味を取り出すためにパターン辞書が構築されており、機械翻訳 [1] や情緒推定 [2] が研究されている。

パターン辞書は、用言述語文を網羅するものとして日本語語彙大系 [3] が既に存在する。重文複文を網羅するパターン辞書に鳥バンク [4] が既に存在する。ここで、単文の名詞述語文を網羅するパターン辞書は存在しない。

このように、既存のパターン辞書で網羅されていない文型が存在するので、その文型を網羅するパターン辞書を作成する必要がある。名詞述語文を網羅するパターン辞書が存在しない理由は、名詞の数が、用言の数と比べ、非常に多く、かつ、名詞そのものが新しく作られるためと思われる。日本語語彙大系の用言述語文のパターンは、用言を文字に残し、単語ごとにパターンを作成している。用言は約 6,000 語あり、単語に多義性があるので、約 16,000 パターン収集された。名詞述語文のパターンを名詞を文字に残した述語パターンを作ろうとすると、前述した通り非常に多いので、パターンが膨大な数になる上、名詞そのものが新しく作られるため網羅できなくなる問題もある。ゆえに、名詞述語文を対象としたパターンを作成するには、述語の名詞を変数化せざるを得ない。述語を変数化したので、述語の単語ごとに意味を分けることはできないので、名詞述語文の構造が伝える意味を分析する必要がある。

次に、パターン辞書が構築できたと想定しよう。一般に、パターン辞書を用いた意味解析は、入力文をパターンと照合すると、複数の照合結果が得られる。その中から照合結果の選択を行うことで、意味が解析されたことになる。日本語語彙大系では、意味属性コードを利用し選択処理を行なっている。名詞述語文の意味解析においても、意味属性コードを利用した選択処理を試みる。さらに、意味属性コードだけでは、選択できないことが予想されるため、新たな方法についても試みる。

1.2 関連研究

1.2.1 日本語語彙大系

用言述語文を対象とするパターン辞書として日本語語彙大系 [3] が存在する。日本語語彙大系は、日英機械翻訳における意味解析を実現するための意味辞書である。「この意味辞書は、単語や表現構造の意味を体系的に分類した意味属性体系、単語に関する知識を収録した単語意味辞書、用言を核とした表現構造を収録した構文意味辞書の3つから構成される。」[3]

日本語語彙大系のパターンは、字面、変数、および、記号で構成され、日本語パターンと英語パターンが対になっている。日本語パターンと照合すると、対になる英語パターンに変換された結果、日英機械翻訳が行われる。変数は、品詞情報を基に単語を表す変数である。 N は名詞を表す変数である。また、変数には「意味属性体系」を用いて、変数の照合を限定させることで照合結果の絞り込みを行っている。

1.2.2 鳥バンク

重文複文を対象とするパターン辞書として鳥バンク [4] が存在する。鳥バンクは、重文複文の表現を意味的に網羅するような表現意味辞書を実現することを目標としたパターン辞書である。鳥バンクは、機械翻訳に用いられている。

鳥バンクのパターンは、字面、変数、および、記号で構成される。字面は、パターンに文字を記載したもので、照合の際は、字面部分は完全一致のみ可能である。変数は、品詞情報を基に単語、句、節を表す変数である。 N は名詞を表す変数である。記号は、汎用性を持たせるために使用する。「/」は、任意の単語列に適合する記号であり「離散記号」と呼ぶ。「|」は、記号の前後の字面や変数を選言的に照合する。大括弧は、括弧内の字面や変数の有無を任意化して照合する。

1.2.3 名詞述語文の分類

益岡らは、名詞述語を『動詞と形容詞が単独で述語になるのに対して、名詞は単独では述語になれない。名詞と結合して述語を作るのが「判定詞」である』とした [5]。そして、名詞述語の機能は「述部が指し示す集合体に主部が属する場合(例1)」、「主部と述部が同一のものを指し示す場合(例2)」、および「主部と述部の間に直接的な論理関係が存在しない場合(例3)」の3つがあるとしている。

例 1 源氏物語 は 平安時代の作品 だ。

例 2 紫式部 は 源氏物語の作者 だ。

例 3 私 は 紫式部 だ。

例 2 は述部を集合と考えると，要素がたったひとつの集合と考えることができる．よって本研究では，例 1 と例 2 の名詞述語の機能は同様とする．

今田は，名詞述語文を主語と述語の意味関係から，「主語の帰属する範疇を述語で述べる文（範疇叙述型）」，「主語の価格や数量などの属性を述語で述べる文（属性叙述型）」，「内包的概念を表す主語の外延を述語で述べる文（外延叙述型）」，および「主語と述語に意味関係が認められない文（その他）」の 4 つに分類できるとした [6]．今田の分類は，益岡らの分類と比べ，名詞述語文を網羅している範囲が広いので，今田の分類を基に名詞述語文を扱う．今田の分類の詳細は第 2 章に記述する．

1.3 目的

本研究では，以下の項目を目的とする．

- 今田の分類 [6] に基づき，名詞述語文を対象とするパターン辞書を作成する．
- 照合結果の選択の方法を試みる．
 - － 意味属性コードを利用する方法
 - － 機械学習を利用する方法

以上に向けて，第 2 章では名詞述語文の分類を示す．第 3 章では，今田の例文，および，Wikipedia の文を対象に文の構造について分析を行う．特に分析における注意点を示す．第 4 章では分析結果をまとめてパターンの作成方法を示す．第 5 章では，入力文の意味を解析するためのパターン辞書の運用方法を示す．第 6 章ではパターンによる解析性能を実験により評価する．第 7 章では考察を述べる．最後に第 8 章でまとめを述べる．

第2章 今田による名詞述語文の分類

生産物名詞が主語となる名詞述語文は、主語と述語との意味関係に基づき、以下の4つに分類される [6] .

2.1 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。

例4 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。

例5 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。

例6 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。

例7 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。

例 4, 5 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、例 4 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、例 5 のように比較的上位の類概念を表す語が現れる場合、さらには例 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。例えば例 4 の「ドキュメンタリー」は博物学的分類の一種であるが、例 7 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。例 7 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。例 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、例 5, 6 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、例 5 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、述語の連体修飾表現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

2.2 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。

例 8 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。

例 9 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。

[6] より引用

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．

一般的にいえば，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，述語の連体修飾表現は値（以降は「属性値」と呼称）となるが，述語が属性および属性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（[6] より引用）

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 ([6] より引用)

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.

2.3 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。

例 10 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。

例 11 掲げる 旗 は 民主主義 だ。

例 12 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。

[6] より引用

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例 12 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、述語の連体修飾表現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

2.4 その他

範疇叙述型，属性叙述型，および，外延叙述型のように，主語と述語の間に意味関係が認められない場合は，その他に分類する．今田は以下の例文を示した．

例 13 時計 は既に ロスタイム だった。

[6] より引用

名詞述語文の主語と述語の関係は，帰属関係，対象と属性の関係，内包と外延の関係などの一般的な意味論的关系に即して解釈されるが，そのような関係が認められない場合には，両者の関係は様々な語用論的原則に従って解釈される．そのような事例の代表的なものは，ウナギ文という名称でよく知られている．どのような文をウナギ文と認定するかは難しい問題であるが，主語と述語の関係を一般的な意味論的关系に即して類型化することができない名詞述語文の事例としては，次のような文を挙げることができる．例 13 では「ロスタイム」は「時計」の進み具合を表している．

本研究では，「その他」は研究対象に含めない．

第3章 名詞述語文の分析

3.1 目的

パターンの作成，および，運用を行う際，文の構造を把握するために，事例の分析を行う．

3.2 項目

今田の例文，および，Wikipedia の文を分析対象とする．Wikipedia は 2014 年 11 月 4 日時点のデータを使用する．本章では主に以下の点を明らかにする．

- 今田の 3 分類の是非
- 主語および述語についての抽出範囲
- はが構文の扱い
- 定形表現について

3.3 今田の 3 分類の是非

複雑な文を分析していると，判断に迷うことがある．また，範疇叙述型は属性の「種類」や「分野」などの属性叙述型，外延叙述型は属性が「外延」の属性叙述型と，「3 つの分類を全て属性叙述型として扱う」という考え方ができる．そこで，分類に対する考え方をまとめておく．

- 外延叙述型と属性叙述型は主語の外延を述べるか属性を述べるかの違いがあり別ものである．
- 範疇叙述型は属性叙述型に含まれるという指摘があるが，主語と述語の相対性を見ているので別のものとする．

- － 属性叙述型は性質や程度などを表すもの、または（拡大解釈すると）種類などを表すものなどがある。

例 14（程度）このカレーは300円だ。

例 15（種類）電子レンジは電化製品だ。

例 14 では、述語は主語の価格を表しており、主語と述語に意味的な相対性はない。例 15 では、述語は主語の種類を表しており、主語と述語に包含関係があるため、主語と述語に意味的な相対性があるといえる。

- － 文中の主語や述語の抽象度は相対的に見るので、主語のみまたは述語のみから抽象度を確定させることはできない。
- － ゆえに、拡大解釈をせずに、例 15 は、範疇叙述型と考える。

以上より、本研究では、分類は今田の3分類とすることに問題ないとする。また、範疇叙述型を属性叙述型と外延叙述型の補集合と考えることにする。

3.4 主語についての抽出範囲

文中の主部のうちどの範囲までをパターン処理上の主語として抽出すれば良いかについて考え方を示す。

例 16 初期の 自動車 は手作りのものである。

例 16' 初期の自動車 は手作りのものである。

例 17 これを1本のケーブルで接続できるように端子を1つにまとめたもの がD端子である。

もし、主語の抽出範囲を例 16 の下線部のみとするならば、主語名詞だけでは具体的な実体ではなく、一般的なものを指すため、不適切である。これに対し、もし、述部と対応の取れる程度に具体的になるように例 16' の下線部を主語の抽出範囲とするならば、主語の抽出範囲は連体修飾表現までとなる。しかし、例 16' の基準を例 17 に用いると下線部のとおりの主語の抽出範囲となり、連体修飾節が接続詞や複数の動詞を含むという長い表現となるのだが、連体修飾節をパターンで全てをマッチさせても、型名の判定で全て使うとは限らないので、パターンで処理する上では不要である。したがって、「修飾表現が主語の前にあるか」という判断までを必要としつつ、例 16 の下線部を主語の抽出範囲と定める。

3.5 述語についての抽出範囲

主語と同様に，パターン処理上の述語としての抽出範囲の考え方を示す．

例 18 蒸気機関車は蒸気機関で動く 機関車 である。

述語の抽出範囲は，主語と同様の考えで，例 18 の下線部を述語の抽出範囲とし，述語の修飾語句を述語とは別の意味のある表現部分（後述の追加情報）と考える．

3.6 「はが構文」の扱い

「はが構文」は文の分析において判断を誤りやすい文であるので注意を述べる．以下に，はが構文，および，非はが構文（「*」を表記）の例を示す．

例 19 強度部材に用いられる材料は鋼鉄が主流である。

例 20* そのパスポートは私が落としたものだ。

「は格」の前の単語（例 19 では「材料」），および「が格」の前の単語（例 19 では「鋼鉄」）の係り先が同一のものを「はが構文」という．例 20* のように「は格」の前の単語，および「が格」の前の単語の係り先が別であれば「はが構文」ではないことに注意を要する（図 3.1）．

一方，はが構文は（広い意味での）主語と述語の関係が 2 つ存在することにも注意を要する．

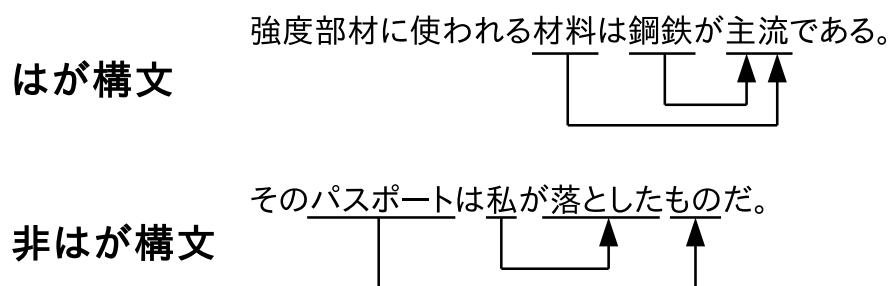


図 3.1: 「はが構文」と「非はが構文」の例

3.7 定形表現について

定形表現を含む文についても判断を誤りやすいため注意を述べる．

例 21 ブルーチーズは、チーズの一種である。

「～の一種」という表現がある。「の一種」は、その前の単語が主語の範疇を表すと判定する手がかりとなる表現である．これを本研究では定形表現と呼ぶことにする．定形表現を含む文に対しては、定形表現の前の単語に対して主語との関係を比較する．

他の定形表現を Wikipedia から収集を試みた．使用データは、「～である」または「～だ」の表現の前の単語を抽出し、頻度の高い単語から順に人手で分析を行った．上位 250 件までで、8 種類を得た．分析を行った単語を一部載せる（表 3.1）．定形表現を効率的に発見することは、今後の課題とする．

得た表現 一種，1 種，1 つ，一つ，ひとつ，仲間，総称，一分野

表 3.1: Wikipedia 頻度上位単語

頻度順序	単語字面
1	もの
2	一つ
3	可能
4	データ
5	キャラクター
6	郡
7	道
8	一般的
9	番組
10	意味
20	女優
30	バス
40	局
50	オーケストラ
60	学校
70	こと
80	不明
90	企業
100	一種
150	施設
200	歌手
250	通り

第4章 パターンの作成

3章で、名詞述語文の分析を行うことにより、今田の分類で問題ないこと、および、名詞述語文での注意すべき点が確認できた。これを基にパターンの作成を行う。

4.1 パターン化基準

パターンの作成にあたって、3章の調査を基にパターン化基準を作成する。これにより、パターン作成者個人に依存する揺れを抑える。全ての型名で共通する基準と、各型名での基準があるので、共通(G)、属性叙述型(A)、外延叙述型(E)、範疇叙述型(C)の順で記述する(表4.1)。パターン化基準は、作業者に対する指示であり、「確認する」こと、「記述する」こと、および「注意する」ことが混在している。

パターン作成の成果物はパターン辞書である。パターン辞書は多数のエントリで構成する。1つのエントリは、「原文(1文以上)」、「パターン」、「選択条件」、および「応用情報」で構成される。

原文は、パターン、選択条件、および、応用情報を作成する際に参照する文である。

パターンは、1.2.2節を参照。本研究では、変数に、*AN*はアルファベットと数字の連続、*MT*は連体修飾表現、*MD*は判定詞を含むモダリティ表現を表す変数を追加している。

選択条件は、照合結果の選択を行う際に使用する条件で、「修飾表現が主語の前にあるか」というフラグである(3.4節参照)。また、変数の意味属性制約もここに該当する。

応用情報とは、パターンを照合した際、出力される情報であり、意味解析結果として得たい情報のテンプレートである。名詞述語文の場合、名詞述語文の叙述の型名、および、3つ組がある。3つ組の情報は、上位語、下位語、追加情報、実体、属性、および、属性値であり、この組み合わせを3つ組とする。追加情報は、述語の前にある連体修飾表現である(3.5節参照)。

表 4.1: パターン化基準

基準 ID	パターン化基準
G1	はが構文と非はが構文の 2 種類が存在する .
G2	主語はパターンでは名詞 (連続した名詞の場合一つの名詞とみなす) と前後の接辞まで含め変数化する .
G3	主語の前に連体修飾表現がある場合 , 変数化し , 選択条件にフラグ T を記す . また , 連体修飾表現の変数の前に離散記号を記す .
G4	主語の前に連体修飾表現がない場合 , 選択条件にフラグ F を記す . また , 主語の変数の前に離散記号を記す .
G5	はが構文の場合 , 助詞は , 1 つ目は「は」で 2 つ目が「が」で , 助詞の直後の読点まで字面に残す .
G6	非はが構文の場合 , 助詞は「は」または「が」で , 助詞の直後の読点まで字面に残す . また , 「は」および「が」の選択化 , 読点の任意化を行う .
G7	述語はパターンでは名詞 (連続した名詞の場合一つの名詞とみなす) と前後の接辞まで含め変数化する .
G8	述語の前に連体修飾表現がある場合 , それが追加情報で , 変数化し , そうでない場合 , 追加情報は付けない . また , 助詞と追加情報の間の単語を削除し , 離散記号を記す .
G9	文末の判定詞を変数化し , 以降を削除する .
G10	主語または述語の具体性が低い名詞と判断した場合は , 該当名詞を字面に残したパターンを別に作成し , 計 2 パターン用意する .
G11	型名を確認し , 応用情報に記す .
A1	属性叙述型から取り出される情報は , 実体 , 属性 , 属性値の 3 つであり , 型名および 3 つ組を応用情報とする .
A2	主語は実体を表す .
A3	追加情報および述語は事例によってそれぞれが属性または属性値になり得る .
A4	取り出される情報の述語は , 抽象的な意味を持つ単語である .
E1	外延叙述型から取り出される情報は , 上位語 , 下位語 , 追加情報の 3 つであり , 型名および 3 つ組を応用情報とする .
E2	主語は上位語を表す .
E3	述語は下位語を表す .
E4	相対的に見て主語が上位語で , 述語が下位語である .
E5	述語が固有名詞で , 主語は固有名詞でない場合は外延叙述型といえる .
E6	主語が形式名詞のように抽象度が高い場合は , 外延叙述型といえる .
C1	範疇叙述型から取り出される情報は , 下位語 , 上位語 , 追加情報の 3 つであり , 型名および 3 つ組を応用情報とする .
C2	主語は下位語を表す .
C3	述語は上位語を表す .
C4	相対的に見て主語が下位語で , 述語が上位語である .
C5	主語が固有名詞で , 述語は固有名詞でない場合は範疇叙述型といえる .
C6	主語の前に指示詞がある場合は , 主語の抽象度が低いとみなすことができる .
C7	判定詞の前に単語自体で文を範疇叙述型と判定できる定形表現がある場合は , 該当部分を字面に残し , 直前の名詞と接辞までを述語のように扱う .

4.2 作成の様子

『「しめ縄」は、悪霊の侵入を防ぐ道具だった。』をパターン原文としてエントリを表4.1の順に従って作成する例を示す。

4.2.1 パターンの作成

パターン化基準に従い、パターンを作成する。

G1 はが構文と非はが構文の2種類が存在する。

はが構文か非はが構文かの確認を行う。例文を見るとはが構文ではないことが確認できる。

G2 主語はパターンでは名詞（連続した名詞の場合一つの名詞とみなす）と前後の接辞まで含め変数化する。

例文の主語は『「しめ縄」』なので、該当部分を変数 N に変更する。また、変数には他の変数との区別のために番号をつける。

作業後:

$N1$ は、悪霊の侵入を防ぐ道具だった。

G3 主語の前に連体修飾表現がある場合、変数化し、選択条件にフラグ T を記す。また、連体修飾表現の変数の前に離散記号を記す。

G4 主語の前に連体修飾表現がない場合、選択条件にフラグ F を記す。また、主語の変数の前に離散記号を記す。

例文に主語の前に連体修飾表現はないので、主語の変数の前に離散記号を付記する。

作業後:

$/N1$ は、悪霊の侵入を防ぐ道具だった。

G5 はが構文の場合、助詞は、1つ目は「は」で2つ目が「が」で、助詞の直後の読点まで字面に残す。

G6 非はが構文の場合、助詞は、「は」または「が」で、助詞の直後の読点まで字面に残す。また、「は」および「が」の選択化、読点の任意化を行う。

例文は非はが構文なので、助詞は「は」で、直後に読点を確認できる。助詞を「が」にも対応できるように選択化を行い、読点の任意化を行う。

作業後:

$/N1(\text{は} | \text{が})[\text{,}]$ 悪霊の侵入を防ぐ道具だった。

G7 述語はパターンでは名詞（連続した名詞の場合一つの名詞とみなす）と前後の接辞まで含め変数化する。

例文の述語は「道具」なので、該当部分を変数 N に変更する。

作業後:

$/N1(\text{は} | \text{が})[\text{,}]$ 悪霊の侵入を防ぐ $N2$ だった。

G8 述語の前に連体修飾表現がある場合、それが追加情報で、変数化し、そうでない場合、追加情報は付けない。また、助詞と追加情報の間の単語を削除し、離散記号を記す。

例文には述語の前に連体修飾表現があるので、該当部分の「悪霊の侵入を防ぐ」を変数 MT にし、変数の前に離散記号を付記する。

作業後:

$/N1(\text{は} | \text{が})[\text{,}]/MT2 N3$ だった。

G9 文末の判定詞を変数化し、以降を削除する。

例文の文末の判定詞「だった」を変数 MD にし、判定詞以降を削除する。

作業後:

$/N1(\text{は} | \text{が})[\text{,}]/MT2 N3 MD4$

G10 主語または述語の具体性が低い名詞と判断した場合は、該当名詞を字面に残したパターンを別に作成し、計2パターン用意する。

例文の述語「もの」は具体性が低い名詞と判断できる。よって、述語の字面を残したパターンを別に作成する（今回は字面を残したパターンについては以降の作業では割愛する）。

作業後:

$/N1(\text{は} | \text{が})[\text{,}]/MT2 N3 MD4$

$/N1(\text{は} | \text{が})[\text{,}]/MT2 \text{もの} MD3$

G11 型名を確認し、応用情報に記す。

例文の主語は『「しめ縄」』で述語は「もの」である。主語と述語の意味関係を見ると主語の『「しめ縄」』が具体性が高い。よって、例文は範疇叙述型である。次は範疇叙述型 (C) のパターン化基準に移る。

- C1 範疇叙述型から取り出される情報は、下位語、上位語、追加情報の3つであり、型名および3つ組を応用情報とする。
- C2 主語は下位語を表す。
- C3 述語は上位語を表す。
- C4 相対的に見て主語が下位語で、述語が上位語である。
- C5 主語が固有名詞で、述語は固有名詞でない場合は範疇叙述型といえる。
- C6 主語の前に指示詞がある場合は、主語の抽象度が低いとみなすことができる。
- C7 判定詞の前に単語自体で文を範疇叙述型と判定できる定形表現がある場合は、該当部分を字面に残し、直前の名詞と接辞までを述語のように扱う。

例文に判定詞の前に定形表現はないので作業を終了する。

4.2.2 エントリの作成

パターン化基準に従い、選択条件、および、応用情報に情報を記述する。

- G3 主語の前に連体修飾表現がある場合、変数化し、選択条件にフラグ T を記す。また、連体修飾表現の変数の前に離散記号を記す。
- G4 主語の前に連体修飾表現がない場合、選択条件にフラグ F を記す。また、主語の変数の前に離散記号を記す。

例文に主語の前に連体修飾表現はないので、選択条件の「主語前」に F を付記する。

- G11 型名を確認し、応用情報に記す。

例文の主語は『「しめ縄」』で述語は「もの」である。主語と述語の意味関係を見ると主語の『「しめ縄」』が具体性が高い。よって、例文は範疇叙述型なので応用情報の型名に記す。また、次は範疇叙述型のパターン化基準に移る。

- C1 範疇叙述型から取り出される情報は、下位語、上位語、追加情報の3つであり、型名および3つ組を応用情報とする。

C2 主語は下位語を表す．

C3 述語は上位語を表す．

範疇叙述型の応用情報の3つ組の組み合わせおよび対応する変数を記載する．

C4 相対的に見て主語が下位語で，述語が上位語である．

C5 主語が固有名詞で，述語は固有名詞でない場合は範疇叙述型といえる．

C6 主語の前に指示詞がある場合は，主語の抽象度が低いとみなすことができる．

C4～C6の項目より相対的に見て例文の主語が下位語で，述語が上位語になっているかを再度確認する．

C7 判定詞の前に単語自体で文を範疇叙述型と判定できる定形表現がある場合は，該当部分を字面に残し，直前の名詞と接辞までを述語のように扱う．

以上の作業を行うことにより，エントリが作成できる．作成したパターンは，主語前の連体修飾表現のないパターンで，主語の『「しめ縄」』が変数 $N1$ と対応している．助詞「は」を「が」と選択化し，読点を任意化する．述語の前の連体修飾表現「悪霊の侵入を防ぐ」は $MT2$ に対応している．述語「道具」は $N3$ に対応している．末尾の判定詞が $MD4$ である．主語と述語の意味関係を見ると，『「しめ縄」』と「道具」で述語の「道具」が具体性が低く，上位語と言える．これに対する主語は，相対的に下位語となる．主語と述語の関係から，型名は範疇叙述型となる．

エントリ例

原文

原文：「しめ縄」は、悪霊の侵入を防ぐ道具だった。

パターン

パターン： / $N1$ (は | が)[,]/ $MT2$ $N3$ $MD4$

選択条件

主語前： F

応用情報

型名： 範疇叙述型

下位語： $N1$

上位語： $N3$

追加情報： $MT2$

4.3 作成結果

文献 [6] に示された例文 33 文，および，2014 年 11 月 4 日時点の Wikipedia 「自動車」ページから得た具体物が主語の名詞述語文 8 文の計 41 文を基にパターンを作成したところ，21 エントリを得た．次節以降に分類毎のエントリを示す．型名ごとの内訳を表 4.2 に示す．なお，本研究では，名詞述語に対する主語が存在し，かつ，末尾に判定詞が存在する名詞述語文を対象とする．

表 4.2: エントリの内訳

型名	エントリ数
範疇叙述型	10
属性叙述型	6
外延叙述型	5
合計	21

4.3.1 範疇叙述型

エン트리1のパターンは、主語の前の連体修飾表現があり、述語の前の連体修飾表現もある範疇叙述型のパターンである。原文が6つからエントリが構成される。原文1を例にとると、主語の前の連体修飾表現の「トルコの」は、変数 $MT1$ で、主語の『「汗と絨毯」』は $N2$ に対応している。助詞の「は」を「が」と選択処理ができるようにし、読点を任意化する。述語の前の連体修飾表現の「職人たちを追った」は $MT3$ で、述語の「ドキュメンタリー」が $N4$ に対応する。判定詞は $MD5$ である。主語と述語の意味関係は、主語は『「汗と絨毯」』で、述語が「ドキュメンタリー」である。『「汗と絨毯」』は「ドキュメンタリー」つまり記録映像のタイトルだとわかる。『「汗と絨毯」』は数ある「ドキュメンタリー」の中のひとつを指している。ゆえに、述語は上位語で、主語が下位語である。

エン트리1

原文

- 原文1: トルコの「汗と絨毯」は、職人たちを追ったドキュメンタリーだ。
原文2: 冷蔵庫にある製品は国内向けのものだった。
原文3: 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だ。
原文4: 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だ。
原文5: 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。
原文6: 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だ。

パターン

パターン: / $MT1$ $N2$ (は | が)[,]/ $MT3$ $N4$ $MD5$

選択条件

主語前: T

応用情報

型名: 範疇叙述型
下位語: $N2$
上位語: $N4$
追加情報: $MT3$

エントリ2のパターンは、エントリ1とほぼ同様だが、述語が「もの」と共通して具体性が低いので、述語を字面に残した範疇叙述型のパターンである。原文は5つで構成する。原文1を例にすると、主語の前の連体修飾表現の「冷蔵庫にある」は、変数 $MT1$ で、主語の「製品」は $N2$ に対応している。助詞の「は」を「が」と選択処理をできるようにし、直後に読点はないが、任意化する。述語の前の連体修飾表現「国内向けの」を $MT3$ に対応し、述語の「もの」は具体性が低いので字面に残す。判定詞は $MD4$ である。主語と述語の意味関係は、どちらも具体性が低いが、述語の「もの」は形式名詞とみなせるので、述語のほうが具体性が低い。ゆえに、主語が下位語で、述語が上位語である。

エントリ2

原文

原文 1: 冷蔵庫にある製品は国内向けのものだった。

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だ。

原文 3: 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だ。

原文 4: 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。

原文 5: 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だ。

パターン

パターン: $/MT1 N2(は | が)[,]/MT3 もの MD4$

選択条件

主語前: T

応用情報

型名: 範疇叙述型

下位語: $N2$

上位語: もの

追加情報: $MT3$

エントリ3のパターンは、主語の前の連体修飾表現、述語の前の連体修飾表現ともに存在しない範疇叙述型のパターンである。原文は1つで構成する。原文を例にすると、主語の「発泡酒」が変数 $N1$ に対応する。助詞の「は」を「が」に選択処理できるようにし、読点はないが読点の任意化をする。述語の「アルコール飲料」が $N2$ に対応し、判定詞は $MD3$ である。主語と述語の意味関係は、「発泡酒」は「アルコール飲料」の中のひとつであるので、主語が下位語で、述語が上位語の関係である。

エントリ3	
原文	
原文: 発泡酒はアルコール飲料だ。	
パターン	
パターン: / $N1$ (は が)[,]/ $N2$ $MD3$	
選択条件	応用情報
主語前: F	型名: 範疇叙述型
	下位語: $N1$
	上位語: $N2$
	追加情報: null

エントリ 4 のパターンは、主語の前の連体修飾表現はなく、述語の前の連体修飾表現はある範疇叙述型のパターンである。原文は 12 で構成する。原文 1 を例にすると、主語の『「しめ縄」』が変数 $N1$ と対応している。助詞「は」を「が」と選択化し、読点を任意化する。述語の前の連体修飾表現「悪霊の侵入を防ぐ」は $MT2$ に、述語「道具」は $N3$ に対応している。末尾の判定詞が $MD4$ である。主語と述語の意味関係を見ると、『「しめ縄」』は「道具」の中のひとつであるので、主語は、下位語で、述語は上位語の関係である。

エントリ 4

原文

- 原文 1: 「しめ縄」は、悪霊の侵入を防ぐ道具だった。
 原文 2: ワゴン車はレンタル会社から借り出されたものだ。
 原文 3: ラップは若者へのメッセージだ。
 原文 4: 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。
 原文 5: 石油はわが国の主権のシンボルだ。
 原文 6: 銃は同士官学校から盗み出したものだった。
 原文 7: バッジは、実行グループの一人が落としたものだ。
 原文 8: 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。
 原文 9: 伏見城は太閤秀吉の城だった。
 原文 10: 抗がん剤が自社の製品だった。
 原文 11: フレーム形式は独立した骨格部材に車室を構成する外殻構造が載せられたもので、古くから自動車の車体構造として用いられ、現在でも貨物車を中心に採用されている。
 原文 12: モノコック形式は車室を構成する外殻自体が強度部材として作られた構造で、20世紀半ば頃から自動車の車体構造として普及しはじめて、現在の乗用車のほとんどで採用されている。

パターン

パターン: / $N1$ (は | が)[、]/ $MT2$ $N3$ $MD4$

選択条件

主語前: F

応用情報

型名: 範疇叙述型

下位語: $N1$

上位語: $N3$

追加情報: $MT2$

エントリ5のパターンは、エントリ4とほぼ同様に主語の前の連体修飾表現がなく、述語の前の連体修飾表現のある範疇叙述型のパターンだが、述語が共通して「もの」と具体性が低いので、述語を字面に残す。原文は5つで構成する。原文1を例にすると、主語の「ワゴン車」は変数 $N1$ に対応する。助詞の「は」を「が」に選択処理ができるようにし、読点はないが読点の任意化をする。述語の前の連体修飾表現の「レンタル会社から借り出された」は $MT2$ に対応し、述語の「もの」は具体性が低いので字面に残す。判定詞は $MD3$ である。主語と述語の意味関係は、述語の「もの」は具体性が低いと判断できるので、主語は下位語で、述語が上位語の関係である。

エントリ5

原文

- 原文 1: ワゴン車はレンタル会社から借り出されたものだ。
 原文 2: 銃は同士官学校から盗み出したものだった。
 原文 3: バッジは、実行グループの一人が落としたものだ。
 原文 4: 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。
 原文 5: フレーム形式は独立した骨格部材に車室を構成する外殻構造が載せられたもので、古くから自動車の車体構造として用いられ、現在でも貨物車を中心に採用されている。

パターン

パターン: / $N1$ (は | が)[、]/ $MT2$ もの $MD3$

選択条件

主語前: F

応用情報

型名: 範疇叙述型
 下位語: $N1$
 上位語: もの
 追加情報: $MT2$

エントリ 6 のパターンは、エントリ 4 と同様に、主語の前の連体修飾表現はなく、述語の前の連体修飾表現はある範疇叙述型のパターンであるが、述語の「道具」が具体性が低いと判断できるので、述語を字面に残したパターンも作成する。原文は 1 つで構成する。原文を例にすると、主語の『「しめ縄」』が変数 $N1$ と対応している。助詞「は」を「が」と選択化し、読点を任意化する。述語の前の連体修飾表現「悪霊の侵入を防ぐ」は $MT2$ に、述語「道具」は具体性が低いと判断できるので字面に残す。末尾の判定詞が $MD4$ である。主語と述語の意味関係を見ると、『「しめ縄」』は「道具」の中のひとつであるので、主語は、下位語で、述語は上位語の関係である。

エントリ 6	
原文	
原文: 「しめ縄」は、悪霊の侵入を防ぐ道具だった。	
パターン	
パターン: / $N1$ (は が)[、]/ $MT2$ 道具 $MD3$	
選択条件	応用情報
主語前: F	型名: 範疇叙述型 下位語: $N1$ 上位語: 道具 追加情報: $MT2$

エントリ7のパターンは、エントリ4と同様に、主語の前の連体修飾表現はなく、述語の前の連体修飾表現はある範疇叙述型のパターンであるが、述語の「メッセージ」を具体性が低いと判断したので、述語を字面に残したパターンも作成する。原文は1つで構成する。原文を例にすると、主語の「ラップ」が変数 $N1$ と対応している。助詞「は」を「が」と選択化し、読点を任意化する。述語の前の連体修飾表現「若者への」は $MT2$ に、述語「メッセージ」は具体性が低いと判断したので字面に残す。末尾の判定詞が $MD4$ である。主語と述語の意味関係を見ると、「メッセージ」を具体性が低いのと判断したので主語の「ラップ」は、下位語で、述語の「メッセージ」は上位語の関係とする。

エントリ7	
原文	
原文: ラップは若者へのメッセージだ。	
パターン	
パターン: / $N1$ (は が)[、]/ $MT2$ メッセージ $MD3$	
選択条件	応用情報
主語前: F	型名: 範疇叙述型 下位語: $N1$ 上位語: メッセージ 追加情報: $MT2$

エントリ 8 のパターンは、エントリ 4 と同様に、主語の前の連体修飾表現はなく、述語の前の連体修飾表現はある範疇叙述型のパターンであるが、述語の「シンボル」を具体性が低いと判断したので、述語を字面に残したパターンも作成する。原文は 1 つで構成する。原文を例にすると、主語の「石油」が変数 *N1* と対応している。助詞「は」を「が」と選択化し、読点を任意化する。述語の前の連体修飾表現「わが国の主権の」は *MT2* に、述語の「シンボル」は具体性が低いと判断したので字面に残す。末尾の判定詞が *MD4* である。主語と述語の意味関係を見ると、「シンボル」を具体性が低いのと判断したので主語の「石油」は、下位語で、述語の「シンボル」は上位語の関係とする。

エントリ 8	
原文	
原文: 石油はわが国の主権のシンボルだ。	
パターン	
パターン: / <i>N1</i> (は が)[、]/ <i>MT2</i> シンボル <i>MD3</i>	
選択条件	応用情報
主語前: F	型名: 範疇叙述型 下位語: <i>N1</i> 上位語: シンボル 追加情報: <i>MT2</i>

エントリ9のパターンは、エントリ4と同様に、主語の前の連体修飾表現はなく、述語の前の連体修飾表現はある範疇叙述型のパターンであるが、述語の「製品」を具体性が低いと判断できるので、述語を字面に残したパターンも作成する。原文は1つで構成する。原文を例にすると、主語の「抗がん剤」が変数 $N1$ と対応している。助詞「が」を「は」と選択化し、読点を任意化する。述語の前の連体修飾表現「自社の」は $MT2$ に、述語の「製品」は具体性が低いと判断できるので字面に残す。末尾の判定詞が $MD4$ である。主語と述語の意味関係を見ると、「抗がん剤」はある「製品」の中のひとつという関係なので、主語の「抗がん剤」は、下位語で、述語の「製品」は上位語の関係とする。

エントリ9	
原文 原文: 抗がん剤が自社の製品だった。	
パターン パターン: / $N1$ (は が)[、]/ $MT2$ 製品 $MD3$	
選択条件 主語前: F	応用情報 型名: 範疇叙述型 下位語: $N1$ 上位語: 製品 追加情報: $MT2$

エントリ 10 のパターンは、エントリ 4 と同様に、主語の前の連体修飾表現はなく、述語の前の連体修飾表現はある範疇叙述型のパターンであるが、述語の「構造」を具体性が低いと判断したので、述語を字面に残したパターンも作成する。原文は 1 つで構成する。原文を例にすると、主語の「モノコック形」が変数 $N1$ と対応している。助詞の「は」を「が」と選択化し、読点を任意化する。この助詞の後に助詞の「が」が存在するが、助詞「は」の前の単語と係り先が違うので例文ははが構文ではない。述語の前の連体修飾表現の「強度部材として作られた」は $MT2$ に、述語の「構造」は具体性が低いと判断したので字面に残す。末尾の判定詞が $MD4$ である。主語と述語の意味関係を見ると、「モノコック形式」はある「構造」の中のひとつという関係なので、主語の「モノコック形式」は、下位語で、述語の「構造」は上位語の関係とする。

エントリ 10	
原文	
原文: モノコック形式は車室を構成する外殻自体が強度部材として作られた構造で、20世紀半ば頃から自動車の車体構造として普及しはじめて、現在の乗用車のほとんどで採用されている。	
パターン	
パターン: / $N1$ (は が)[、]/ $MT2$ 構造 $MD3$	
選択条件	応用情報
主語前: F	型名: 範疇叙述型 下位語: $N1$ 上位語: 構造 追加情報: $MT2$

4.3.2 属性叙述型

エントリ 11 のパターンは、主語の前の連体修飾表現がなく、述語の前の連体修飾表現がある属性叙述型のパターンである。原文は 6 つで構成する。原文 1 を例にすると、主語の「つぶしあん」が変数 $N1$ に対応する。助詞の「は」を「が」に選択処理できるようにし、直後に読点がないが読点を任意化する。述語の前の連体修飾表現の「抑えた」は $MT2$ で、述語の「甘さ」は $N3$ に対応する。判定詞は $MD4$ である。主語と述語の意味関係は、述語の「甘さ」に注目すると、味覚を表す属性表現である。よって、原文 1 は属性叙述型である。述語の前の連体修飾表現を注目すると、「抑えた」と程度を表している。今回の属性は「甘さ」で、属性値を述語の前の連体修飾表現と述語の「抑えた甘さ」とした。

エントリ 11

原文

原文 1: つぶしあんは抑えた甘さだ。
原文 2: 偽造投票用紙は本物と同じ薄い緑色だ。
原文 3: ステージはたばこの火などでは燃えにくい材質だ。
原文 4: VTR は 12・0%増の 51万7000台だ。
原文 5: ビールが免許行政下で一物一価だった。
原文 6: 自動車は使用者に多くの便益を与えるが、反面、生命・健康・安全などの市民の基本的権利を侵害する存在である。

パターン

パターン: / $N1$ (は | が)[、]/ $MT2$ $N3$ $MD4$

選択条件

主語前: F

応用情報

型名: 属性叙述型
実体: $N1$
属性: $N3$
属性値: $MT2$ $N3$

エントリ 12 のパターンは、主語の前の連体修飾表現、述語の前の連体修飾表現ともに存在する属性叙述型のパターンである。原文は 4 つで構成する。原文 1 を例にすると、主語の前の連体修飾表現の「前に垂れた」が変数 $MT1$ で、主語の「帯」が $N2$ に対応する。助詞の「は」を「が」に選択処理できるようにし、直後に読点はないが読点を任意化する。述語の前の連体修飾表現の「鯉の滝昇りの」は $MT3$ で、述語の「図」が $N4$ に対応する。判定詞は $MD5$ である。主語と述語の意味関係は、述語に注目すると、「図」であり、広い意味を持つが、述語の前の連体修飾表現から「鯉の滝昇り」を描画しているという属性表現とわかる。主語は「帯」なので上位下位の関係は成り立たない。属性は主語の「図」で、属性値は述語の前の連体修飾表現と述語の「鯉の滝昇りの図」とした。

エントリ 12

原文

原文 1: 前に垂れた帯は鯉の滝昇りの図だ。

原文 2: 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だ。

原文 3: 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。

原文 4: 初期の自動車は手作りであるため非常に高価なものであり、貴族や大金持ちだけが所有できるものであった。

パターン

パターン: $/MT1 N2(\text{は} | \text{が})[\text{,}]/MT3 N4 MD5$

選択条件

主語前: T

応用情報

型名: 属性叙述型

実体: $N2$

属性: $N4$

属性値: $MT3 N4$

エントリ 13 のパターンは、エントリ 12 と同様に主語の前の連体修飾表現、述語の前の連体修飾表現ともに存在する属性叙述型のパターンであるが、述語の「図」は広い意味を持つため具体性が低いと判断し、述語を字面に残すパターンを作成する。原文は 1 つで構成する。原文を例にすると、主語の前の連体修飾表現の「前に垂れた」が変数 *MT1* で、主語の「帯」が *N2* に対応する。助詞の「は」を「が」に選択処理できるようにし、直後に読点はないが読点を任意化する。述語の前の連体修飾表現の「鯉の滝昇りの」は *MT3* に対応し、述語の「図」を字面に残す。判定詞は *MD4* である。主語と述語の意味関係は、述語に注目すると、「図」であり、広い意味を持つが、述語の前の連体修飾表現から「鯉の滝昇り」を描画しているという属性表現とわかる。主語は「帯」なので上位下位の関係は成り立たない。属性は主語の「図」で、属性値は述語の前の連体修飾表現と述語の「鯉の滝昇りの図」とした。

エントリ 13	
原文	
原文: 前に垂れた帯は鯉の滝昇りの図だ。	
パターン	
パターン: / <i>MT1</i> <i>N2</i> (は が)[,]/ <i>MT3</i> 図 <i>MD4</i>	
選択条件	応用情報
主語前: T	型名: 属性叙述型 実体: <i>N2</i> 属性: 図 属性値: <i>MT3</i> 図

エントリ 14 のパターンは、主語の前の連体修飾表現、述語の前の連体修飾表現ともに存在しない属性叙述型のパターンである。原文は3つで構成する。原文 1 を例にすると、主語の「タンク」が変数 $N1$ に対応する。助詞の「は」を「が」に選択処理できるようにし、直後に読点はないが読点を任意化する。述語の「グラスファイバー製」が $N2$ に対応する。判定詞は $MD3$ である。主語と述語の意味関係は、述語を注目すると「～製」と材料を表す属性表現である。述語の前の連体修飾表現はないので、属性、属性値ともに「グラスファイバー製」を指す。

エントリ 14	
原文	
原文 1: タンクはグラスファイバー製だ。 原文 2: カレーは三百円だ。 原文 3: 実験器具は実費だ。	
パターン	
パターン: $/N1(\text{は} \text{が})[,]/N2 MD3$	
選択条件	応用情報
主語前: F	型名: 属性叙述型 実体: $N1$ 属性: $N2$ 属性値: $N2$

エントリ 15 のパターンは、は格の主語の前に連体修飾表現のあるはが構文のパターンで、型名は属性叙述型である。原文は1つで構成する。原文を例にすると、は格の主語の前の連体修飾表現「車体の強度部材に用いられる」が変数 $MT1$ で、主語の「材料」が $N2$ に対応する。助詞の「は」を字面に残し、は格の主語の前の連体修飾表現である「鋼鉄」が $N3$ に対応する。直後の助詞の「が」を字面に残す。述語の「主流」を $N4$ に対応する。判定詞は $MD5$ である。主語と述語の意味関係は、述語を注目すると、「主流」は、ここでは「大多数」などの意味を持つ属性表現である。属性は「主流」で、属性値は「鋼鉄」を指す。

エントリ 15

原文

原文: 車体の強度部材に用いられる材料は鋼鉄が主流で、近年ではアルミニウム合金や炭素繊維強化プラスチックなどの複合材料を用いたものも市販されるようになってきている。

パターン

パターン: / $MT1$ $N2$ は/ $N3$ が $N4$ $MD5$

選択条件

主語前: T

応用情報

型名: 属性叙述型

実体: $N2$

属性: $N4$

属性値: $N3$

エントリ 16 のパターンは、はが構文であり、は格の主語の前の連体修飾表現、が格の主語の前の連体修飾表現ともに存在する属性叙述型のパターンである。原文は 2 つで構成する。原文 1 を例とすると、は格の主語の前の連体修飾表現「運転者の」が変数 $MT1$ で、は格の主語「座席」が $N2$ に対応する。直後の助詞の「は」を字面に残す。が格の主語の前の連体修飾表現「椅子形の」が $MT3$ で、が格の主語「もの」が $N4$ に対応する。直後の助詞の「が」を字面に残す。述語の「主流」が $N5$ に対応する。判定詞は $MD6$ である。主語と述語の意味関係は、述語を注目すると、「主流」は、ここでは「大多数」などの意味を持つ属性表現である。属性は「主流」で、属性値は「椅子形のもの」を指す。

エントリ 16

原文

原文 1: 運転者の座席は座部と背もたれを備えた椅子形のものが主流である。

原文 2: 駐車ブレーキを操作する装置は、レバーを引き上げる方式のものが主流であるが、古いトラックやワンボックスカーでは杖状のレバーを車体前方の奥から手前に引き寄せる方式のものもある。

パターン

パターン: $/MT1 N2$ は [、] $/MT3 N4$ が $N5 MD6$

選択条件

主語前: T

応用情報

型名: 属性叙述型

実体: $N2$

属性: $N5$

属性値: $MT3 N4$

4.3.3 外延叙述型

エントリ 17 のパターンは、主語の前の連体修飾表現、述語の前の連体修飾表現ともに存在する外延叙述型のパターンである。原文は2つで構成する。原文 1 を例にすると、主語の前の連体修飾表現を「陽性の」が変数 $MT1$ で、主語の「食物」が $N2$ に対応する。助詞の「は」を「が」に選択処理できるようにし、直後の読点はないが読点を任意化する。述語の前の連体修飾表現の「地面の下で育つ」が $MT3$ で、述語の「野菜、海草、大豆、小魚」が $N4$ に対応する。判定詞は $MD5$ である。主語と述語の意味関係は、ある「食物」の外延を「野菜、海草、大豆、小魚」と述べている文である。よって、主語が上位語で、述語が下位語である。

エントリ 17

原文

原文 1: 陽性の食物は地面の下で育つ野菜、海草、大豆、小魚だ。

原文 2: 最初の自動車は蒸気機関で動く蒸気自動車で、1769年にフランス陸軍の技術大尉ニコラ＝ジョゼフ・キュニョーが製作したキュニョーの砲車であると言われている。

パターン

パターン: $/MT1 N2(\text{は} | \text{が})[,]/MT3 N4 MD5$

選択条件

主語前: T

応用情報

型名: 外延叙述型

上位語: $N2$

下位語: $N4$

追加情報: $MT3$

エントリ 18 のパターンは、主語の前の連体修飾表現、述語の前の連体修飾表現ともに存在する外延叙述型のパターンである。原文は1つで構成する。原文を例にすると、主語の前の連体修飾表現を「宇宙研の」が変数 $MT1$ で、主語の「主力ロケット」が $N2$ に対応する。助詞の「が」を「は」に選択処理できるようにし、直後の読点はないが読点を任意化する。述語の前の連体修飾表現の「今回の」が $MT3$ で、述語の「M3S2」は、アルファベットと数字の並びの型番なので $AN4$ に対応する。判定詞は $MD5$ である。主語と述語の意味関係は、ある「主力ロケット」の外延を「M3S2」と述べている文である。よって、主語が上位語で、述語が下位語である。

エントリ 18	
原文	
原文: 宇宙研の主力ロケットが今回のM3S2だ。	
パターン	
パターン: $/MT1 N2(は が)[,]/MT3 AN4 MD5$	
選択条件	応用情報
主語前: T	型名: 外延叙述型
	上位語: $N2$
	下位語: $AN4$
	追加情報: $MT3$

エントリ 19 のパターンは、主語の前の連体修飾表現は存在し、述語の前の連体修飾表現はない外延叙述型のパターンである。原文は2つで構成する。原文1を例にすると、主語の前の連体修飾表現の「ともに使われた」が変数 $MT1$ で、主語の「凶器」は $N2$ に対応する。助詞の「は」を「が」に選択処理できるようにし、直後の読点を任意化する。述語の「けん銃」が $N3$ に対応する。判定詞は $MD4$ である。主語と述語の意味関係は、ある「凶器」の外延を「けん銃」と述べている文である。よって、主語が上位語で、述語が下位語である。

エントリ 19	
原文	
原文 1: ともに使われた凶器は、けん銃だった。 原文 2: 掲げる旗は民主主義だ。	
パターン	
パターン: $/MT1 N2(は が)[,]/N3 MD4$	
選択条件	応用情報
主語前: T	型名: 外延叙述型 上位語: $N2$ 下位語: $N3$ 追加情報: null

エントリ 20 のパターンは、主語の前の連体修飾表現、述語の前の連体修飾表現ともに存在しない外延叙述型のパターンである。原文はひとつで構成する。原文を例にすると、主語の「商品」が変数 $N1$ に対応する。助詞の「は」を「が」に選択処理できるようにし、直後に読点はないが読点を任意化する。述語の「一年物スーパー定期」が $N2$ に対応する。判定詞は $MD3$ である。主語と述語の意味関係は、ある「商品」の外延を「一年物スーパー定期」と述べている文である。よって、主語が上位語で、述語が下位語である。

エントリ 20	
原文	
原文: 商品は一年物スーパー定期だ。	
パターン	
パターン: $/N1(\text{は} \text{が})[\text{,}] / N2 MD3$	
選択条件	応用情報
主語前: F	型名: 外延叙述型
	上位語: $N1$
	下位語: $N2$
	追加情報: null

エントリ 21 のパターンは、エントリ 20 と同様に主語の前の連体修飾表現、述語の前の連体修飾表現ともに存在しない外延叙述型のパターンであるが、主語の「商品」は具体性が低いと判断できる。原文は1つで構成する。原文を例にすると、主語の「商品」を字面に残す。助詞の「は」を「が」に選択処理できるようにし、直後に読点はないが読点を任意化する。述語の「一年物スーパー定期」が変数 $N1$ に対応する。判定詞は $MD2$ である。主語と述語の意味関係は、エントリ 20 と同様に、ある「商品」の外延を「一年物スーパー定期」と述べている文である。よって、主語が上位語で、述語が下位語である。

エントリ 21	
原文	
原文: 商品は一年物スーパー定期だ。	
パターン	
パターン: /商品 (は が)[,]/ $N1$ $MD2$	
選択条件	応用情報
主語前: F	型名: 外延叙述型 上位語: 商品 下位語: $N1$ 追加情報: null

第5章 パターンの運用

5.1 運用手順

まず，入力文を全てのパターンと照合すると，照合結果が複数得られる．照合結果には応用情報および変数へのバインド値が含まれている．1つのパターン（エントリ）においても，バインドが複数通りになりえるので，照合結果は複数通りになる．そこで，次に，照合結果の選択を行う．最後に，採用した照合結果の応用情報（バインド済み）を，入力文の意味解析結果として出力する．

5.2 実装

パターン照合に用いるの辞書は，4.3節のエントリで構成される．パターン作成に用いた原文を実験に用いるため，原文については，オープン性はない．パターン照合のツールは，MorphAnalyzer（形態素解析）および Patlap（本研究室で作成されたパターン処理ツール）を用いる．

5.3 運用種類

照合結果の選択は2つの手法を利用する．詳細は次節以降に記載する．

手法 1: 意味属性を利用した照合結果の選択

手法 2: 機械学習を利用した照合結果の選択

5.4 意味属性を利用した照合結果の選択

意味属性を利用した照合結果の選択は，以下の手順で行う．

手順 1: 追加情報および属性値のある適合結果を優先して選択

手順 2：変数がより具体的な適合結果を優先して選択

手順 3：主語，述語の上下関係をもとに優先して選択

手順 4：原文との意味的な近さをもとに選択

ここで，手順 1，2 は，文法レベルでパターンが適切にマッチしたかを調べ，手順 3，4 は，意味レベルでパターンが適切にマッチしたかを調べるものである．

手順 1 は，それぞれの型の文の構造を確認する．範疇叙述型は述語の前，外延叙述型は主語の前に連体修飾表現があるかを確認を行う．属性叙述型では，属性値が抽出されているか確認する．手順 2 では，変数がより具体的な適合結果を選択することにより，文法的な具体性を確認する．手順 3 は，範疇叙述型および外延叙述型の意味的な文の構造を確認する．主語および述語に付与されている意味コードにより，範疇叙述型では主語が下位で述語が上位の関係，外延叙述型では主語が上位で述語が下位の関係が，それぞれ成り立っている適合結果を選択する．手順 4 は 5.4.1 節で述べる．

5.4.1 原文との意味的な近さ

入力文とパターンの原文が意味的に近いことを条件に，最適な適合結果を選ぶ．上位語や属性表現の意味の近さを重視しており，式 (5.1) で最適な適合結果 \hat{m} を求める．

$$\hat{m} = \arg \min_{\substack{m \in M \\ (C_{\text{主}}, C_{\text{述}}) \leftarrow m \\ c_1 \in C_{i_{\text{主}}}, c_2 \in C_{\text{主}} \\ c_3 \in C_{i_{\text{述}}}, c_4 \in C_{\text{述}}}} \{w_1 d(c_1, c_2) + w_2 d(c_3, c_4)\} \quad (5.1)$$

ここで， M は適合結果の集合， m は適合結果（原文，パターン，型名，意味制約，3 つ組，および，パターンマッチで代入される変数値で構成．なお， M は原文の数および代入のバリエーションにより増大）， $(C_{\text{主}}, C_{\text{述}})$ は m からとり出される意味制約であり， $C_{\text{述}}$ はなかでも述語に対する集合（例 1 では， $\{1035, 893\}$ ）， $C_{i_{\text{主}}}$ は入力文 i の主語の意味コードの集合（一語につき複数のコードが存在）， $C_{\text{主}}$ ， $C_{i_{\text{述}}}$ も同様の集合， (w_1, w_2) は， m の型が外延叙述型ならば $(1.0, 0.1)$ とし，その他ならば $(0.1, 1.0)$ とする． d はシソーラス [3] において，2 つの意味コードの距離を表す関数である．

シソーラスは，名詞の概念の上下関係を木構造で表す．ルートノードは最上位であり，抽象度が最も高い．下位ノードであるほど具体的である．ノードは意味コードが付与さ

れている．アークは上下関係のあることを表す．関数 d は，2つの意味コードに対応するノード間のアーク数を返すことで，意味の距離を表す．

5.4.2 実行の様子

「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。」を入力文とした「意味属性を利用した照合結果の選択」を利用する運用例を示す．

Step1 パターン照合

例文を全てのパターンと照合を行う．その結果として照合結果が複数得られる．以下に，照合結果を示す．

入力文：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。

照合結果 1

パターン原文

原文 1: 前に垂れた帯は鯉の滝昇りの図だ。

原文 2: 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だ。

原文 3: 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。

原文 4: 初期の自動車は手作りであるため非常に高価なものであり、貴族や大金持ちだけが所有できるものであった。

マッチしたパターン

パターン: `/MT1 N2(は | が)[,]/MT3 N4 MD5`

選択条件

主語前: T

応用情報

型名: 属性叙述型

実体: 道路

属性: 幹線

属性値: 生活用品や復旧資材を運ぶ貴重な幹線

照合結果 2

パターン原文

原文 1: トルコの「汗と絨毯」は、職人たちを追ったドキュメンタリーだ。

原文 2: 冷蔵庫にある製品は国内向けのものだった。

原文 3: 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だ。

原文 4: 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だ。

原文 5: 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。

原文 6: 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だ。

マッチしたパターン

パターン: /MT1 N2(は | が)[,]/MT3 N4 MD5

選択条件

主語前: T

応用情報

型名: 範疇叙述型

下位語: 道路

上位語: 幹線

追加情報: 生活用品や復旧資材を運ぶ貴重な

照合結果 3

パターン原文

原文 1: 陽性の食物は地面の下で育つ野菜、海草、大豆、小魚だ。

原文 2: 最初の自動車は蒸気機関で動く蒸気自動車で、1769年にフランス陸軍の技術大尉ニコラ = ジョゼフ・キュニョーが製作したキュニョーの砲車であると言われている。

マッチしたパターン

パターン: /MT1 N2(は | が)[、]/MT3 N4 MD5

選択条件

主語前: T

応用情報

型名: 外延叙述型

上位語: 道路

下位語: 幹線

追加情報: 生活用品や復旧資材を運ぶ貴重な

Step2 照合結果の選択

意味属性を利用した照合結果の選択の手順の処理を行う。

手順 1: 追加情報および属性値のある適合結果を優先して採用

例文の全ての照合結果で、追加情報または属性値が存在するので全ての候補を採用する。

手順 2: 変数がより具体的な適合結果を優先して採用

マッチしたパターンの変数は、全てのマッチしたパターンが同様なので全て採用する。

手順 3: 主語、述語の上下関係をもとに優先して採用

例文の主語の意味属性コードは(418 道路)で、述語の意味属性コードは(417 交通路, 2669 経路)である。この内で、主語の(418 道路)が下位、述語の(417 交通路)が上位の関係あたる。よって、主語が下位語で、述語が上位語の範疇叙述型の照合結果 2 を採用する。

手順 4: 原文との意味的な近さをもとに採用

マッチしたパターンの原文の中から (5.1) 式で最も意味的に近い原文からスコアを得る。パターン辞書だと、入力文が含まれているので最も近い原文は入力文でスコアは 0.0 である。同一の文を覗いた場合は、「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だ。」で、スコアは 12.0 となる。この選択を行う段階で、照合結果 2 だけなので、これを最も原文に近い候補として採用する。

以上の処理で、例文は以下の照合結果を採用する。

入力文: 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。

採用した照合結果

パターン原文

- 原文 1: トルコの「汗と絨毯」は、職人たちを追ったドキュメンタリーだ。
- 原文 2: 冷蔵庫にある製品は国内向けのものだった。
- 原文 3: 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だ。
- 原文 4: 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だ。
- 原文 5: 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。
- 原文 6: 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だ。

マッチしたパターン

パターン: /MT1 N2(は | が)[,]/MT3 N4 MD5

選択条件

主語前: T

応用情報

型名: 範疇叙述型

下位語: 道路

上位語: 幹線

追加情報: 生活用品や復旧資材を運ぶ貴重な

5.5 機械学習を利用した照合結果の選択

表 4.1 のパターン化基準を参考に作成した素性生成ルール（表 5.1）で照合結果毎に真理値ベクトルを作成する（図 5.1）。このベクトルを利用し，機械学習で各照合結果を採用すべきか破棄すべきかの識別を行う。

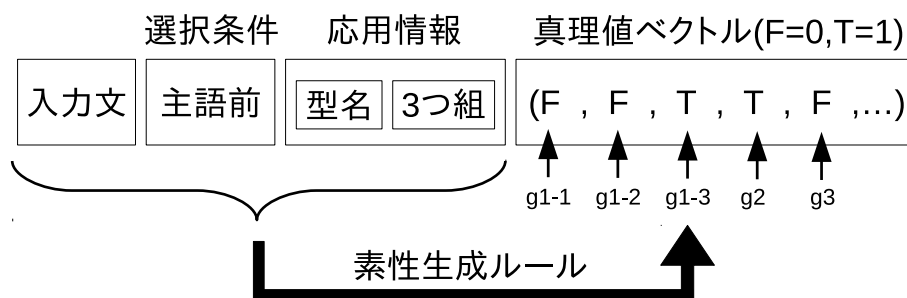


図 5.1: 真理値ベクトルのイメージ

5.5.1 使用ツール

機械学習で利用したツールは，TinySVM0.09（SVM ツール）[8] である。はが構文の判定は，CaboCha（係り受け解析器）[9] を利用し，「(広い意味での) 主語 2 つの係り先が，両方とも述語になっているか」で判定する。意味属性コード，品詞，および，記号の判定は MorphAnalyzer の結果を利用する。

5.5.2 実行の様子

5.4.2 節と同様に「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。」を入力文とした「機械学習を利用した照合結果の選択」を利用する運用例を示す。

Step1 パターン照合

5.4.2 節と同様なため割愛

Step2 照合結果の選択

まず，表 5.1 の素性生成ルールから，照合結果の真理値ベクトルを作成する。次に，機械学習で各照合結果を採用すべきか破棄すべきかの識別を行う。

手順 1: 照合結果の真理値ベクトルを作成

表 5.1: 素性生成ルール

素性 ID	素性生成ルール (T/F)
g1-1	パターンの原文は、はが構文か
g1-2	マッチした文は、はが構文か
g1-3	g1-1 および g1-2 の判定は一致しているか
g2	主語はあるか
g3	パターン原文の主語の前に連体修飾節があるか
g4-1	マッチした文の主語の前に連体修飾節があるか
g4-2	g3 および g4-1 の判定は一致しているか
g5	はが構文かつ、助詞「は」および「が」があるか
g6	非はが構文かつ、助詞「は」または「が」があるか
g7	述語はあるか
g8-1	パターン原文に追加情報はあるか
g8-2	マッチした文に追加情報はあるか
g8-3	g8-1 および g8-2 の判定は一致しているか
g9	文末に判定詞はあるか
g10	マッチしたパターンに名詞の字面が残っているか
g11	照合結果で想定されている型名は何か (非真理値)
a4-1	属性の意味属性コードは全て抽象か
a4-2	属性の意味属性コードは全て具体か
a4-3	属性の接尾辞は“数え方の辞典”[7]にあるか
a4-4	属性の品詞は用言性名詞か
a4-5	属性の品詞は形容動詞転成型名詞か
ec4-1	意味属性コード上で主語は述語の上位か (不明は F)
ec4-2	意味属性コード上で主語は述語の下位か (不明は F)
ec4-3	主語と述語の共通する意味属性コードはあるか
ec4-4	主語に意味属性コードはあるか
ec4-5	述語に意味属性コードはあるか
e5-1	述語は固有名詞か
e5-2	述語はカギ括弧で囲われているか
e5-3	述語はアルファベットの単語か
e6	主語は形式名詞か
e7	主語の前に連体修飾表現があり、追加情報がないか
c5-1	主語は固有名詞か
c5-2	主語はカギ括弧で囲われているか
c5-3	主語はアルファベットの単語か
c6	主語の直前に指示詞があるか
c7	マッチしたパターンに定形表現があるか

手順 2: 機械学習で各照合結果を識別

手順 1 で作成した真理値ベクトルを入力として、学習データは今田の例文（33 文，正解データ 33 件），および，2014 年 11 月 4 日時点の Wikipedia の「自動車」ページの名詞述語文（8 文，正解データ 11 件）の 2 種類のデータ両方とし，同一の原文以外のデータを使用する．学習データも入力と同様に真理値ベクトル化して，機械学習で識別を行う．

機械学習で各照合結果を識別の結果以下の照合結果が採用される．

入力文: 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。

採用した照合結果

パターン原文

- 原文 1: トルコの「汗と絨毯」は、職人たちを追ったドキュメンタリーだ。
原文 2: 冷蔵庫にある製品は国内向けのものだった。
原文 3: 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だ。
原文 4: 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だ。
原文 5: 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。
原文 6: 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だ。

マッチしたパターン

パターン: /MT1 N2(は | が)[,]/MT3 N4 MD5

選択条件

主語前: T

応用情報

型名: 範疇叙述型
下位語: 道路
上位語: 幹線
追加情報: 生活用品や復旧資材を運ぶ貴重な

第6章 実験

6.1 実験の目的

本研究で作成したパターン辞書，および，照合結果の選択における2つの手法の解析性能の評価を行う．

6.2 実験の条件

クローズドテスト，クロスバリデーション，および，オープンテストを行う．使用する文は以下のとおりである（表6.1，6.2，6.3）．

- 今田の例文（33文，正解データ33件）
- 2014年11月4日時点のWikipediaの「自動車」ページの名詞述語文（8文，正解データ11件）
- 2014年11月4日時点のWikipediaの「潜水艦」ページの名詞述語文（26文，正解データ27件）

表 6.1: 型の内訳（今田の例文）

型名	文数
範疇叙述型	17
属性叙述型	11
外延叙述型	5
合計	33

表 6.2: 型の内訳（「自動車」の文）

型名	文数
範疇叙述型	5
属性叙述型	4
外延叙述型	2
合計	11

実験結果の表での手法の表記は，「意味属性を利用した照合結果の選択」は「意味属性を利用」，「機械学習を利用した照合結果の選択」は「機械学習を利用」と表記する．一致数は，型名および3つ組が正解データと全て同一の場合にカウントする． $T = \langle \text{フィルタリングの対象数} \rangle$ であるが，照合結果数であるので，入力文数よりも多い．現在行なっ

表 6.3: 型の内訳 (「潜水艦」の文)

型名	文数
範疇叙述型	11
属性叙述型	12
外延叙述型	4
合計	27

ている照合結果の選択は，入力文 1 文に対して照合結果を 1 つ選択するのではなく，フィルタリングを行う．ゆえに，再現率 R ((6.1) 式)，適合率 P ((6.2) 式)，および， F 値 ((6.3) 式) を用いて，3 つ組の抽出性能を評価する．

$$R = \frac{\langle \text{一致数} \rangle}{\langle \text{採用すべき数} \rangle} \quad (6.1)$$

$$P = \frac{\langle \text{一致数} \rangle}{\langle \text{採用数} \rangle} \quad (6.2)$$

$$F \text{ 値} = \frac{2PR}{(P + R)} \quad (6.3)$$

6.3 クローズドテスト

クローズドテストは，今田の例文 (33 文，正解データ 33 件)，および，2014 年 11 月 4 日時点の Wikipedia の「自動車」ページの名詞述語文 (8 文，正解データ 11 件) の 2 種類のデータ両方をクローズドテストを行った (表 6.4)．結果は，「意味属性を利用」の方法は，57 件採用し，43 件一致した，「機械学習を利用」の方法は，38 件採用し，27 件一致した．「機械学習を利用」の方法は，採用数は抑えることができているが，一致数が「意味属性を利用」の方法よりも減少している．再現率 R ，適合率 P ，および， F 値全てで「意味属性を利用」の方法が上回った．

表 6.4: 今田の例文および「自動車」文の意味解析の性能 (クローズド)

手法	T	再現率 R	適合率 P	F 値
意味属性を利用	132	0.98 (43/44)	0.75 (43/ 57)	0.85
機械学習を利用	132	0.61 (27/44)	0.71 (27/ 38)	0.66
フィルタなし	132	1.00 (44/44)	0.33 (44/132)	0.50

6.4 クロスバリデーション

クロスバリデーションは、クローズドテストのデータを用いて、leave-one-out cross-validation を行った (表 6.5)。結果は、「意味属性を利用」の方法は、57 件採用し、42 件一致した。「機械学習を利用」の方法は、46 件採用し、27 件一致した。クローズドテストと同様に、「機械学習を利用」の方法は、採用数は抑えることができているが、一致数が「意味属性を利用」の方法よりも減少している。再現率 R 、適合率 P 、および、 F 値全てで「意味属性を利用」の方法が上回った。

表 6.5: 今田の例文および「自動車」文の意味解析の性能 (クロスバリデーション)

手法	T	再現率 R	適合率 P	F 値
意味属性を利用	132	0.95 (42/44)	0.74 (42/ 57)	0.83
機械学習を利用	132	0.61 (27/44)	0.59 (27/ 46)	0.60
フィルタなし	132	1.00 (44/44)	0.33 (44/132)	0.50

6.5 オープンテスト

オープンテストは、2014 年 11 月 4 日時点の Wikipedia の「潜水艦」ページの名詞述語文 (26 文、正解データ 27 件) のオープンテストを行った。機械学習に用いる学習データは、クローズドテストで使用したデータの照合結果 132 件である (表 6.6)。結果は、「意味属性を利用」の方法は、29 件採用し、16 件一致した。「機械学習を利用」の方法は、25 件採用し、15 件一致した。「機械学習を利用」の方法は、「意味属性を利用」の方法と比較し、採用数を 4 件減少させ、一致数を 1 件減少にとどめた。適合率 P 、および、 F 値が、「機械学習を利用」の方法をわずかに上回った。再現率 R は「意味属性を利用」の方法がわずかに上回った。

表 6.6: 「潜水艦」文の意味解析の性能 (オープン)

手法	T	再現率 R	適合率 P	F 値
意味属性を利用	127	0.59 (16/27)	0.55 (16/ 29)	0.57
機械学習を利用	127	0.56 (15/27)	0.60 (15/ 25)	0.58
フィルタなし	127	0.96 (26/27)	0.20 (26/127)	0.33

第7章 考察

7.1 2つの手法の比較

表 6.4, 6.5 で, 再現率 R , 適合率 P , および, F 値の全てで「意味属性を利用した照合結果の選択」の性能が高いことが確認できる. これより, パターンの原文と入力文を意味属性で比較することは, 照合結果の選択において有効だと考える. 表 6.6 では, ほとんど差は確認できなかった. オープンテストの結果を見ると, 傾向として, 網羅性は「意味属性を利用した照合結果の選択」が, 照合結果の絞り込みの性能は「機械学習を利用した照合結果の選択」が有効と確認できる.

7.2 作成したパターン辞書

作成したパターン辞書が入力文とのパターン照合をした時点の一致数を見ると, 表 6.4, 6.5 では取りこぼしはなく, オープンテスト (表 6.6) では, 1 件はパターン照合をした時点で正解データと一致しなかった. 1 件一致しなかったが, 名詞述語文の分析を基に作成したパターン化基準は問題なかったと考える.

入力文: 潜水艦は水上艦艇や航空機に比べ、敵の哨戒網や監視網の突破が容易なので、敵勢力下での物資運搬や、特殊部隊揚陸には適役である。

正解データ

型名: 属性叙述型

実体: 潜水艦

属性: 適役

属性値: 敵勢力下での物資運搬や、特殊部隊揚陸には適役

例に挙げた 1 件のみ照合結果に正解が含まれていなかった. 例文中に存在する「には」は「は格」ではなく「に格」であり「はが構文」のような文の構造ではない. 連体修飾表現にマッチする変数 MT は「には」に対応していないので, 正解データの属性値にマッチできない. よって, 対応には, 変数 MT の拡張する必要がある.

7.3 「意味属性を利用」の方法のクローズドテスト

「意味属性を利用」の方法でのクローズドテストで再現率 R が 1.00 にならなかった。誤りになった文を以下に示す。

入力文: 初期の自動車は手作りであるため非常に高価なものであり、貴族や大金持ちだけが所有できるものであった。

正解データ 1	正解データ 2	正解データ 3
型名: 属性叙述型 実体: 自動車 属性: 手作り 属性値: 手作り	型名: 範疇叙述型 下位語: 自動車 上位語: もの 追加情報: 非常に高価な	型名: 範疇叙述型 下位語: 自動車 上位語: もの 追加情報: 貴族や大金持ちだけが所有できる

例の正解データ 1 が選択されなかった。原因として、実装上の問題で、照合結果が採用された場合、選択されなかった照合結果は全て破棄するので、ひとつの入力文で複数の正解があると、採用されず破棄される照合結果が存在する。

7.4 誤り分析

7.4.1 意味解析の成功例

誤り分析を行う比較として、意味解析の成功例を以下に示す。

意味解析の成功例

パターン原文

- 原文 1: 「しめ縄」は、悪霊の侵入を防ぐ道具だった。
原文 2: ワゴン車はレンタル会社から借り出されたものだ。
原文 3: ラップは若者へのメッセージだ。
原文 4: 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。
原文 5: 石油はわが国の主権のシンボルだ。
原文 6: 銃は同士官学校から盗み出したものだった。
原文 7: バッジは、実行グループの一人が落としたものだ。
原文 8: 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。
原文 9: 伏見城は太閤秀吉の城だった。
原文 10: 抗がん剤が自社の製品だった。
原文 11: フレーム形式は独立した骨格部材に車室を構成する外殻構造が載せられたもので、古くから自動車の車体構造として用いられ、現在でも貨物車を中心に採用されている。
原文 12: モノコック形式は車室を構成する外殻自体が強度部材として作られた構造で、20世紀半ば頃から自動車の車体構造として普及しはじめて、現在の乗用車のほとんどで採用されている。

マッチしたパターン

パターン: /N1(は | が)[、]/MT2 N3 MD4

選択条件

主語前: F

応用情報

型名: 範疇叙述型
下位語: 攻撃型潜水艦
上位語: 潜水艦
追加情報: 敵の水上艦艇や潜水艦などの攻撃を任務とする

正解データ

型名: 範疇叙述型
下位語: 攻撃型潜水艦
上位語: 潜水艦
追加情報: 敵の水上艦艇や潜水艦などの攻撃を任務とする

この例では、型名および3つ組が一致しており、正解であることを確認できる。

7.4.2 固有名詞の処理不足

固有名詞の処理不足の例

パターン原文

原文 1: トルコの「汗と絨毯」は、職人たちを追ったドキュメンタリーだ。

原文 2: 冷蔵庫にある製品は国内向けのものであった。

原文 3: 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だ。

原文 4: 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だ。

原文 5: 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。

原文 6: 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だ。

マッチしたパターン

パターン: /MT1 N2(は | が)[,]/MT3 N4 MD5

選択条件

主語前: T

応用情報

型名: 範疇叙述型

下位語: 輸送型潜水艦

上位語: U 1 5 1 型Uボート

追加情報: 第一次大戦期の

正解データ

型名: 外延叙述型

上位語: 輸送型潜水艦

下位語: U 1 5 1 型Uボート

追加情報: 第一次大戦期の

この例では、選択された型名および3つ組の組み合わせが誤っていた。誤り原因は、述語の「U 1 5 1 型Uボート」を固有名詞とみなす処理ができなかったことが挙げられる。形態素解析で単語毎に区切ると「U 1 5 1」、「型」、「U」、「ボート」と分解される。現在の手法では意味属性や品詞を参照する場合は、この中で末尾の「ボート」のみ参照しているため、述語にマッチした単語全てを参照する、または、主語と述語のマッチした単語数で比較する素性を追加することが考えられる。

7.5 今後の対応すべき課題

本研究での名詞述語文は、「名詞述語に対応する主語があり、かつ、末尾に判定詞のある名詞述語文」を実験対象としていた。また、主語の直後の助詞についても「は」および「が」の2種類のみ使用した。今後の研究では実験対象の範囲を広げるので、以下に対応すべき課題を列挙する。

- 判定詞の存在しない文型への対応

例 22 旅客機は、旅客の輸送を目的とした民間用の飛行機

体言止めの文型（例 22）のような、判定詞の存在しない文型へのパターン照合は困難である。理由は、照合を行うとき、判定詞という文の区切りがないので、どの単語が述語になるのか区別ができないためである。例 22 では、本来は「飛行機」が述語だが、パターンに判定詞の変数がないと、「旅客」「輸送」「目的」「民間用」の名詞についても述語としてパターンが照合する。

- 主語の存在しない文型への対応

例 23 輸送する「貨物」が「旅客」になったものである。

例 23 のように述語の「もの」に対応した主語のない文型の主語なし文がある。が格はあるが、連体修飾節内のが格なので『「貨物」』の係り先は述語の「もの」ではない。このときの主語は、前の文脈に記述されている。このような文型に対して、「主語なし文」と判定する方法が必要である。

- 格助詞の「には」「では」「とは」への対応

例 24 LCCは、安く旅行に行く には 最適である。

例 25 座席のクラス には 3種類あり、級から順に「ファーストクラス」「ビジネスクラス」「エコノミークラス」である。

例 24 の「には」は「に格」として使われている。本研究では、変数で例 24 の「には」に対応していなかった。この場合は「に」と置き換えることが可能なので、変数 *MT* に照合できるように変更を行う。また、例 25 の主格としての「には」についても調査を行う必要がある。格助詞の「では」および「とは」についても同様である。

- 助詞の「も」への対応

例 26 旅客機も輸送機の一つである。

助詞の「は」または「が」と違い、「も」は、例 26 のように並立（他のものと同様にこちらも成り立つ）を表すなど、「は」または「が」とは用法に違いがある。本研究では、助詞の「は」および「が」のみ使用したが、文型の網羅性を高めるために助詞の「も」に対応するパターンを作成する必要がある。

- 接尾辞について

パターンの作成では、接尾辞は名詞として変数化した、しかし、述語に接尾辞がある場合は、「属性叙述型」に分類される傾向があった。ゆえに、述語の接尾辞はパターン上に字面として残す方法も考えられる。

第8章 まとめ

本研究では、名詞述語文を分析することにより、名詞述語文の分類を、範疇叙述型、属性叙述型、および、外延叙述型に分類とすることに問題ないことを確認した。そして、分析を基にしたパターン化の基準を示したので、パターン辞書の構築が可能になった。名詞述語文の例文を基に41文から21エントリを得た。意味解析のためのパターンの運用は、実験により、パターンの網羅性は良好であり、照合結果の選択は、現状では意味属性コードを利用する方法、および、機械学習を利用する方法のどちらも適応できることを確認した。

今後の課題は、現状では学習データが少ないのでデータを増加させること、本研究では取り扱わなかった主語なし文や末尾に判定詞のない名詞述語文の対応、および、重文複文全般を網羅している鳥バンク [4] と本研究で作成したパターン辞書との比較である。

謝辞

本研究を進めるにあたり，徳久雅人講師には，始終に渡り研究の進め方や論文の書き方など，細部に渡る御指導を頂きました．ここに深く感謝致します．また，ご多忙の中，本論文をまとめるにあたって木村周平教授に御助言を頂きました．ここに厚くお礼申し上げます．研究を進めるにあたって種々の御助言を頂きました村田真樹教授に心から御礼申し上げます．3年間に渡り，ご指導頂きました村上仁一准教授に心から御礼申し上げます．

参考文献

- [1] 池原悟, 宮崎正弘, 佐良木昌, 池田尚志, 白井諭, 村上仁一, 徳久雅人: “械翻訳のための日英文型パターン記述言語”, 電子情報通信学会術研究報告, 思考と言語研究会. TL2002-48, pp.1-6.
- [2] 田中努, 徳久雅人, 村上仁一, 池原悟: “結合価パターン への情緒生起情報の付与”, 言語処理学会第 10 回年次大会発表論文集, pp.345-348, 2004.
- [3] 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.
- [4] 鳥バンク : <http://unicorn.ike.tottori-u.ac.jp/toribank/>
- [5] 益岡隆志, 田窪行則: “基礎日本語文法”, くろしお出版, pp.24-27, 1989.
- [6] 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, pp.25-48, 2011.
- [7] 飯田朝子, 町田健: “数え方の辞典”, 小学館, 2004.
- [8] TinySVM : <http://chasen.org/~taku/software/TinySVM/>
- [9] CaboCha : <http://code.google.com/p/cabocha/>