

# 修士論文

## 動詞意味類型の曖昧性解消と 評判分析への応用

長岡技術科学大学大学院 工学研究科  
電気電子情報工学専攻  
学籍番号：11312282

岡田 正平

指導教員 山本 和英 准教授

2015年2月12日



# 動詞意味類型の曖昧性解消と評判分析への応用

岡田 正平

## 概要

日本語の用言は動詞と形容詞に分類される。一般に動詞は客観的に観測可能な動作や変化を表すとされており、形容詞は状態や性質、感情などを表すとされている。しかし、実際の動詞/形容詞の分類基準は動詞がその終止形が母音ウを持つ拍で終わる、形容詞が拍「い」で終わるといった形態的なものであり必ずしも意味的な分類とは一致していない。この問題に対し、本研究室での先行研究において用言を意味的に分類する意味類型が提案された。意味類型は動作、変化、感覚・感情、形容の4種類からなり、先行研究によって全ての動詞に対して取り得る意味類型が人手により付与されている。本研究ではその意味類型を実際の自然言語処理タスクに応用可能にすることを目的とし、まずは動詞の意味類型曖昧性解消器の作成を行った。意味類型曖昧性とは同じ動詞が文脈により異なる意味類型を持つ現象であり、その曖昧性解消とは文脈情報から動詞の意味類型を1つに決定することである。本論文では形態素解析結果と既存の言語資源から得られる情報を利用した素性を用いた教師あり機械学習アプローチを用いて解決を行った。その結果、複数の意味類型候補を持つ動詞に対して約70%の精度、全動詞に対して90%以上の精度で正しく意味類型を1つに決定できるようになった。次に、意味類型の概念をサ変動詞にも適用することを考えた。辞書中に意味類型候補が付与されていなくても、すべての意味類型を候補として意味類型曖昧性解消器を用いることで文中のサ変動詞に対して意味類型を1つに決定することも可能であるが、精度向上のためにできるだけ多くのサ変名詞に対して意味類型候補を付与することが望ましい。そこで用言等換言辞書を用いてサ変名詞を換言し、換言後の表現に含まれる用言を手がかりに意味類型の付与を行った。その結果、全サ変名詞の約1/3にあたる3,934語に対して意味類型候補を付与することができた。これらにより、一般動詞とサ変動詞に意味類型を付与できるようになり、実際の自然言語処理タスクへの応用が可能になった。最後に、実応用として評判分析に意味類型を用いることでその有用性を確認した。具体的にはブログ記事中の各文に対して評判表現が含まれているかないかを分類する評判文抽出タスクを設定し、ルールベース手法において品詞情報を用いた場合と意味類型情報を用いた場合とで比較を行った。その結果意味類型情報を用いた場合、品詞情報を用いた場合と比較して分類精度で4.7ポイント、評判文抽出のF値で27.2ポイント上昇し、意味類型の有用性を確認することができた。

# Semantic Type Disambiguation for Japanese and Application for Sentiment Analysis

Shohei Okada

## Abstract

Japanese predicates consist of adjectives and verbs. It is generally assumed that adjectives signify properties or states of people or objects, and while verbs signify observable actions or changes. However, there are verbs that signify properties or emotion. Therefore, part-of-speech (POS) differs from semantic classification for Japanese predicates. On this subject, semantic types, which aim to classify predicates according to their sense, has been proposed in my laboratory's previous study. The semantic types have four types: action, change, emotion, and modification. All verbs have been annotated for all possible types in the previous study. The purpose of this study is to apply the semantic types for actual Natural Language Processing (NLP) task. First, the semantic type disambiguator, that determines verb type from context, was constructed with supervised machine learning approach. Verbs that have ambiguity of sense can be annotated multiple types. In this case, one of the annotated types can be determined from the context. The disambiguator only utilizes information from the result of morphological analysis and existing language resources. It achieved 70% accuracy of disambiguation for newspaper articles using 10-fold cross validation. Because 80% of all verbs have single semantic type, more than 90% of all verbs can be labeled with a correct semantic type. Second, the verbal nouns were annotated for possible types using a paraphrase resource. The verbal nouns are paraphrased and annotated for possible types based on paraphrased expressions. As a result, 3,934 verbal nouns, 1/3 of all verbal nouns, were annotated for possible types. Finally, the semantic types were applied for sentiment analysis as an application. Specifically, the semantic types were compared with POS in the task of classification whether the sentence includes evaluative expression or not. As a result, the classification accuracy improved 4.7 point and F-measure improved 27.7 point by using the semantic types.

# 目次

<b>1</b>	<b>序論</b>	<b>8</b>
1.1	背景	8
1.2	目的	9
1.3	本論文の構成	9
<b>2</b>	<b>関連研究</b>	<b>10</b>
2.1	動詞の分類	10
2.2	語義曖昧性解消	11
2.3	評判分析・感情解析	11
<b>3</b>	<b>中山の意味類型</b>	<b>13</b>
3.1	形状性用言と作用性用言	13
3.2	意味類型の定義	13
3.3	動詞意味類型辞書	14
<b>4</b>	<b>動詞意味類型曖昧性解消</b>	<b>16</b>
4.1	意味類型曖昧性解消器	16
4.2	評価・考察	20
<b>5</b>	<b>サ変名詞に対する意味類型候補付与</b>	<b>22</b>
5.1	用言等換言辞書	22
5.2	サ変名詞への意味類型候補付与手順	23
5.2.1	意味類型候補付与規則	23
5.3	結果	25
<b>6</b>	<b>評判分析への応用</b>	<b>27</b>
6.1	使用したデータ	27
6.2	評価手法	27
6.3	結果・考察	29
6.3.1	誤り解析	29
6.3.2	変化の意味類型が評判分析に与える影響	32

---

7	結論	33
	使用したツール及び言語資源	35
	参考文献	36
	本論文に関する原著論文	38
	付録 A : サ変名詞に付与した意味類型候補	39
	付録 B : 評判分析の実例	47
	付録 C : 本論文に関する原著論文	57

## 表目次

3.1	付与された意味類型の内訳	15
4.1	評価データ中の意味類型の頻度	20
4.2	意味類型曖昧性解消評価実験結果	20
4.3	意味類型曖昧性解消器の評価実験結果内訳	20
5.1	サ変名詞 3,934 語に対する意味類型候補種類数の内訳	25
5.2	換言辞書を用いた自動付与と手動付与の意味類型候補の比較	26
5.3	手動付与と自動付与の結果に重なりがなかったサ変名詞の内訳	26
6.1	品詞情報のみを用いた評判文抽出	29
6.2	意味類型情報を用いた評判文抽出	29
6.3	KNB コーパス中の評判表現に含まれていた動詞「ある」の項	31

## 図目次

3.1	品詞と意味類型の対応	14
4.1	意味類型曖昧性解消器の概要	17
4.2	名詞意味属性	18
4.3	意味類型曖昧性解消器の入出力例	19
4.4	学習データ量の変化に対する意味類型曖昧性解消器の精度の変化	21
5.1	サ変名詞への意味類型候補付与の流れ	24

# 1 序論

## 1.1 背景

日本語における用言は動詞と形容詞に分類される<sup>1</sup>。その分類基準はその終止形が母音ウを持つ拍で終わる（動詞）、または拍「い」で終わる（形容詞）という形態的なものである。また、一般的に動詞は客観的に観測可能な人や物の動きや変化を、形容詞は物事の特徴や性質を表すとされている。しかし形態的な基準に基づく品詞分類は、その品詞に属する語の意味を完全に分離しているわけではない。例えば、動詞の中には「優れる」のような物事の特徴を表す語や「むかつく」のような人の感情を表す語が存在している。したがって、テキストから動作のみを抽出する目的ですべての動詞を抽出すると動作とは無関係な上記のような動詞が含まれてしまう、などの問題が生じることになる。

この問題に対し、本研究室での先行研究において中山ら<sup>17)</sup>は形態上の特徴に関係なく、その意味によって用言を「作用性用言」と「形状性用言」に分類することを提案した。そのうえで、分類の前段階として動詞、形容詞と作用性用言、形状性用言の中間概念と言える4種類の意味類型（動作、変化、感覚・感情、形容）を定義した。また中山らは、用言のうち形容詞は常に意味類型の「形容」のみを取り、動詞はその語義や用法により同じ動詞でも異なる意味類型を取り得るとし、IPA 品詞体系日本語辞書 [1]（以下、IPA 辞書）中の全ての動詞に対して、取り得る意味類型を人手により付与した SNOW D1: 動詞意味類型辞書 [2]（以下、意味類型辞書）を作成し、公開した。

近年、ブログや SNS 等の Consumer Generated Media (CGM) の普及により個人が意見を発信することが容易になっている。これらの情報は、消費者の「生の声」としてマーケティング等の分野において価値のあるものとみなされ、そのテキスト中から動作を表す表現や形容表現を収集する需要が存在している。例えば倉島ら<sup>13)</sup>はブログに記述された人間の経験情報を抽出する手法を提案している。経験は状況・行動・主観の3要素から構成されるとし、これらの要素の中に動作や評価、感情が含まれている。

収集の際の手がかりに品詞を用いている研究は多く存在しており、これらの手法におい

<sup>1</sup>本論文では特に言及しない限り「形容詞」に形容動詞を含めて考える



て、動詞・形容詞という品詞情報の代わりに意味類型情報を用いることで語の意味をより正確に扱うことができるようになる。その他にも、自然言語処理タスクにおいて品詞情報を用いている手法は多数存在し、意味類型を利用することによる波及効果は動詞・形容詞という品詞情報を用いる全ての処理に及ぶと考える。

## 1.2 目的

本研究では中山らが提案した意味類型を、実際の自然言語処理タスクに応用可能にすることを目的とする。意味類型を応用可能にするためには、

- 動詞の意味類型曖昧性解消
- サ変名詞に対する意味類型候補の付与

を行う必要があると考える。

本論文では同じ動詞が文脈により異なる意味類型を持つ現象を動詞の意味類型曖昧性と定義し、文脈情報から動詞の意味類型を1つに決定することを動詞の意味類型曖昧性解消と定義した。この問題に対しては、形態素解析結果と既存の言語資源から得られる情報を利用した素性を用いた教師あり機械学習アプローチを用いて解決を行う。

次に、一般動詞のみに付与された意味類型候補をサ変動詞にも付与する。辞書中に意味類型候補が付与されていなくても、すべての意味類型を候補として意味類型曖昧性解消器を用いることで文中のサ変動詞に対して意味類型を1つに決定することも可能であるが、精度向上のためにできるだけ多くのサ変名詞に対して意味類型候補を付与することが望ましい。そこで本論文では山本ら<sup>18)</sup>の作成した用言等換言辞書 [3] (以下、換言辞書)を用いたサ変名詞への意味類型の付与を行う。

最後に、実応用として評判分析に意味類型を用いることでその有用性を示す。

## 1.3 本論文の構成

まず2章では関連研究について、動詞の分類、語義曖昧性解消、評判分析・感情解析という3つの観点から紹介する。続く3章では、先行研究である中山による意味類型の定義と意味類型辞書について説明する。4章では動詞の意味類型曖昧性と、その解消について述べ、次に5章にてサ変名詞に対する意味類型の付与について述べる。これらの結果を利用し、6章で意味類型を評判分析に応用することで意味類型の有用性を示す。最後に7章にて本論文の結論を述べる。

## 2 関連研究

### 2.1 動詞の分類

動詞の語義に基づく分類としては、英語においては FrameNet<sup>1)</sup> や VerbNet<sup>8)</sup> が用いられ、多くの研究が存在する。例えば、Croce ら<sup>3)</sup> は係り受け関係とカーネル関数に基づく 2 動詞間の類似度を用いて VerbNet や FrameNet に基づいた動詞の分類を行っている。

日本語における動詞の分類としては竹内ら<sup>14)</sup> による意味役割付与システムが挙げられる。意味役割付与システムは入力文を述語項構造解析を行い、その後述語の語義の同定と項の意味役割付与を行う。言語資源を人手で構築することで、深い処理に向けた高級な情報を付与している。また、本研究ではあくまで動詞の主体の動作や状態に着目しているのに対し、竹内らは動作だけでなく動作対象の状態などのより広い現象に着目している。たとえば「コップを水に満たす」という例において、竹内らの意味類型付与システムでは動詞「満たす」を状態変化ありの動詞（水の状態変化させる）としているが、本研究では「満たす」という行為のみに着目し動作の意味類型を持つとする。これにより意味類型情報を利用した場合には、着目する動詞の主辞（「満たす」の例では動作主）を固定することで特定の対象の動作や変化、状態などを扱うことが可能になる。

動詞をその意味により分類することは、日本語教育の上でも意義がある。谷口<sup>15)</sup> によれば、日本語を正しく使いこなす上で動詞の種類についての理解が重要であるとし、その動詞の分類法の一つにアスペクトによる分類を挙げた。これは金田一<sup>16)</sup> が提唱した分類法であり、具体的には第二種の動詞（継続動詞）は意味類型の動作と、第三種の動詞（瞬間動詞）は変化と、第一種（状態動詞）及び第四種（常に「テイル」を伴う動詞）の動詞は形容と対応していると考えられる。アスペクトによる分類であるので意味類型の感覚・感情に対応する分類はない。また、意味類型上では第一種の動詞と第四種の動詞を区別できない。しかしながら、意味類型を用いることで日本語の文に日本語学習者に向けた情報をある程度付与できると考える。

## 2.2 語義曖昧性解消

本研究における意味類型の曖昧性解消は語義を4種類に限定した語義曖昧性解消 (Word Sense Disambiguation, WSD) と見做すことが出来る。WSD は ACL-SIGLEX<sup>1</sup> が開催する SemEval(旧 SENSEVAL) にて共通タスクとして取り組まれている。各タスクは過去の結果から得られ知見やその時代の状況に合わせて設定されている。一般に WSD は訓練データ作成がボトルネックになる問題があり、近年ではそれを解決すべく教師なし学習のアプローチが注目を集めている。Subgraph-based WSD は教師なし学習 WSD の1つであり、既存の語彙知識ベースから語義をノード、語義間の関係をエッジとするグラフを作成し、入力文脈から決定される部分グラフにおいて、PageRank<sup>2)</sup>等の重要性尺度に基づいて語義を決定する。最新の研究としては Manion<sup>4)</sup>らは部分グラフの構築と語義の選択を反復して行い、語義数の少ない語から WSD を行うことで subgraph-based WSD の性能向上を行う方法を提案している。しかしながら、意味類型間には上位・下位のような関係は存在しないため subgraph-based WSD の手法を適用することはできない。そこで、本研究では文脈情報を用いた教師あり学習アプローチを用いて曖昧性解消を行う。

## 2.3 評判分析・感情解析

品詞情報の代わりに意味類型情報を利用することは、語の意味をより正確に扱うことにつながり、様々なタスクにおいて性能向上に寄与できると考える。本研究では意味類型を評判分析に応用することでその有用性を示す。

近年、個人が Web 上に意見を発信することが可能になったことに伴い、テキストを対象とした評判分析・感情解析への関心が高まっている。品詞情報を用いている研究としては以下のような例が挙げられる。Scholz ら<sup>7)</sup>は語ごとに計算されるエントロピーに基づく指標を品詞毎に平均し素性とする手法で、新聞記事中の文に対する肯定/否定/中性(客観)の極性推定で最高65%の精度を達成している。また、小林ら<sup>12)</sup>は意見を(対象, 属性, 評価値)という3つ組で定義、共起パターンに基づく属性・評価値表現の半自動抽出方法を提案しており、抽出時のノイズ軽減を目的としたフィルタリングの1つに品詞情報を用いている。感情解析においては、Nakagawa ら<sup>5)</sup>は依存構造木のノードの極性を表す隠れ変数と、事前極性や極性反転の有無、品詞等の組み合わせを素性として用いた条件付き確率場を用いて文の感情極性を2値分類する手法を提案し、日本語を対象に最高84.7%の精度を達成している。

<sup>1</sup><http://www.clres.com/siglex.html>

品詞情報を用いない手法の例としては、Socher ら<sup>9)</sup>は再帰的ニューラルネットワークをテンソルの導入によって改良した Recursive Neural TensorNetwork によって、文レベルでの感情極性 2 値分類で 85.4%の精度を達成している。この手法では、入力文中の単語を word embedding で表現し入力としている。

評判分析・感情解析の言語資源に関する研究では、Socher らは Pang ら<sup>6)</sup>のコーパスをベースとし、クラウドソーシングを利用することで感情解析における初の完全なラベル付き構文木である Stanford Sentiment Treebank を作成している。このコーパスの出現により、感情解析における否定表現の振る舞いについて定量的な調査 (Zhu ら<sup>11)</sup>) により「否定表現は修飾する句の感情極性を (単純に) 反転する」という直感に従う仮説が否定されるなど、英語の感情解析において有益な知見が得られている。日本語では人手による解析済みブログコーパスである橋本らの KNB コーパス [4] や、高村らの<sup>10)</sup>単語感情極性対応表、小林ら<sup>12)</sup>および東山らの日本語評価極性辞書 [5] などの資源が利用可能である。橋本らの KNB コーパスは人手で解析を行うことにより、品詞を限定せずに評判表現のアノテーションを行っている。本研究では実際に KNB コーパスを用いて評判分析を行い、品詞情報よりも意味類型情報を用いた方が評判表現を網羅性高く扱うことが出来ることを示す。高村らは、隣り合った電子が同方向のスピンを持ちやすいという性質を言語現象に対応させ、与えられた小規模な単語集合に対する極性と、シソーラスやコーパス等から作られる単語ネットワークを用いて全単語に極性を自動付与する手法を提案した。これにより作成された単語感情極性対応表は、形容詞のみならず名詞や動詞にも -1 ~ 1 の感情極性値を付与しているが、単語と感情極性は 1 対 1 対応であり語義の曖昧性に充分対応していない<sup>2</sup>。小林らはテキスト中からの意見抽出において、意見を (対象, 属性, 評価値) という 3 つ組で定義、共起パターンに基づく属性・評価値表現の半自動抽出方法を提案しており、抽出時のノイズ軽減を目的としたフィルタリングの 1 つに品詞情報を用いている。東山らは名詞の評価極性獲得のため、評価極性にセンシティブな選択選好を持つ述語が存在すること (例えば「心がける」はヲ格に「安全運転」などのポジティブな名詞を取りやすい) に着目した機械学習手法を提案している。本研究では名詞の評価極性は、用言の意味を扱う意味類型と組み合わせることで、評判分析において有効に働くことも示す。

<sup>2</sup>語義の曖昧性が全く考慮していないわけではなく、語義によって異なる極性を持つ同一の語はその感情極性値の絶対値が小さくなるを考える

## 3 中山の意味類型

### 3.1 形状性用言と作用性用言

中山らは用言の形態的な分類と意味的な分類が一致しない問題に対して、用言の新しい分類法として「作用性用言」と「形状性用言」を提案し、その分類における問題と課題を検討した。形状性用言と作用性用言は形態的な分類にはこだわらず純粋に意味的な基準に基づいて分類される。例えば、動詞「優れる」は人や物の性質を表すため形状性用言となる。また、動詞「走る」は通常作用性用言であるが、「虫唾が走る」という1つの表現は形状性用言となる。

### 3.2 意味類型の定義

用言を「形状性用言」と「作用性用言」に分類する上で中山らは「動詞については動作動詞や変化動詞など外的に現れる動きや、心の動きなど内的な活動で客観的に捉えることの出来ない表現が含まれている。さらには『優れる』など性質を表すものまで含まれているため、分類の前処理として各動詞が持つ意味をある程度まとめる必要がある。」と述べたうえで、動作、変化、感覚・感情、形容という4種類の意味類型を定義した。

この意味類型は1動詞に対して1つの意味類型が必ず決定するものではなく、複数の意味類型をもつことを許容している。以下に各意味類型の定義を示す。

**動作** 客観的に観測者が捉えることのできる運動で、その運動が終了すると運動前の状態に戻り、結果状態を表さない動詞を意味類型の動作とする。例えば「泳ぐ」や「食べる」などが動作である。

**変化** 主体に現れる運動の結果状態を表す動詞を意味類型の変化とする。主体が意思を持たず結果のみを表し、かつ運動が終了しても運動前の状態には戻らない表現である。例えば「乾く」や「死ぬ」などが変化である。

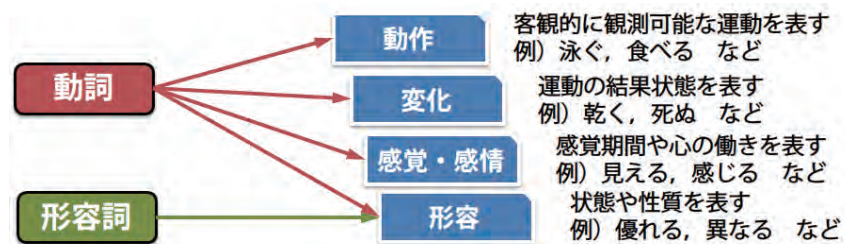


図 3.1 品詞と意味類型の対応

感覚・感情 目, 耳, 皮膚といった感覚器官の活動と, 頭脳や心の働きなどを表す知情意を意味類型の感覚・感情とする。定義では, 意思性のある感覚器官の活動を動作とし, 「見える」や「感じる」など非意志的なものを感覚とするとしている。また「愛する」など, 心の動きだけでなく客観的に観察可能な動きの意味も含んでいる動詞に対して, 感覚・感情の意味類型だけでなく動作も持つと定義している。

形容 人や物の様子や性質, 形, 存在, 関係を表す表現を意味類型の形容とする。形容は一般的な形態上の分類である形容詞の集合と, 意味的な分類の上で同義になると考えられている。

動作または変化の意味類型を持つ動詞は作用性用言に, 形容の意味類型を持つ動詞は形状性用言に漏れ無く対応する。また, 形容詞は常に形容の意味類型のみを持つとする(図 3.1)。

### 3.3 動詞意味類型辞書

中山らは IPA 辞書中の各動詞(12,648 表現)に対して人手で意味類型を付与した。この際, 動詞が持ち得る意味を落とすことの無いよう, 関係すると思われる意味類型全てを付与した。表 3.1 に付与した意味類型の内訳を示す。

表 3.1 付与された意味類型の内訳

動作	変化	感覚・感情	形容	動詞数
	-	-	-	6,637
-		-	-	1,531
-	-		-	1,441
-	-	-		358
		-	-	469
	-		-	1,127
	-	-		190
-			-	396
-		-		113
-	-			72
			-	143
		-		61
	-			35
-				42
				33
			計	12,648

## 4 動詞意味類型曖昧性解消

複数の意味類型候補を持つ動詞は，その文脈によって異なる意味類型を持つ．例えば動詞「満たす」は意味類型辞書中で動作と形容の意味類型候補を持っており，それぞれ次のような文脈で用いられる．

例 1

コップに水を 満たす (動作)

例 2

条件を 満たす (形容)

本研究ではこれを動詞の意味類型曖昧性と定義する．意味類型を実際の自然言語処理タスクに応用するためには，文中の動詞に対してその意味に基づいて意味類型を 1 つに適切に決定しなければならない．本研究では文中の動詞の意味類型を 1 つに決定することを意味類型曖昧性解消と定義し，同じ文中の文脈情報を用いることで動詞の意味類型曖昧性解消が可能になると仮定した．

本章では意味類型曖昧性解消を多値分類問題とみなし，形態素解析結果と既存の言語資源の情報から得られる素性を用いた教師あり機械学習アプローチを用いて意味類型曖昧性解消器を作成する．

### 4.1 意味類型曖昧性解消器

作成した意味類型曖昧性解消器の概要を図 4.1 に示す．まず，曖昧性解消の対象となる動詞を含む入力文を形態素解析器 MeCab[6] により形態素に分割する．この形態素解析結果を元に，以下に示すの 5 種類の素性を生成する．

ただし数式中における各記号は，

$s = \{w_1, \dots, w_m\}$  : 入力文

$w_t$  : 入力文中の曖昧性解消対象動詞 ( $1 \leq t \leq m$ )

$\phi(s, t) = [\phi_x(s, t)]$  : 素性ベクトル



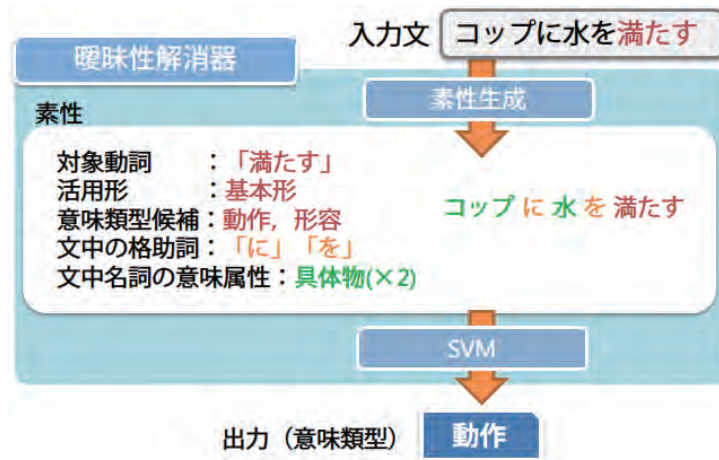


図 4.1 意味類型曖昧性解消器の概要

を表すものとする。

対象動詞 曖昧性解消の対象となる動詞自身を素性とする。意味類型辞書中で複数の意味類型が付与された 2,681 動詞について、その動詞が対象動詞であれば 1、そうでない場合には 0 の 2 値をとる。

$$\phi_{v_i}(s, t) = \begin{cases} 1 & \text{if } w_t = v_i \\ 0 & \text{otherwise} \end{cases} \quad (4.1)$$

ただし、 $v_i \in V$  ( $1 \leq i \leq |V|$ ,  $|V| = 2,681$ ) であり、 $V$  は意味類型辞書中で複数の意味類型が付与された動詞集合を表す。

対象動詞の活用形 動詞の活用形はその動詞の文中における役割に関連すると考え、対象動詞の活用形を素性とする。IPA 辞書中の 19 種類の活用形について、対象動詞の活用形がその活用形であれば 1、そうでない場合には 0 の 2 値をとる。

$$\phi_{f_i}(s, t) = \begin{cases} 1 & \text{if } conj(w_t) = f_i \\ 0 & \text{otherwise} \end{cases} \quad (4.2)$$

ただし、 $conj(w)$  は語  $w$  の活用形、 $f_i \in F$  ( $1 \leq i \leq |F|$ ,  $|F| = 19$ ) であり、 $F$  は IPA 辞書中の全活用形を表す。

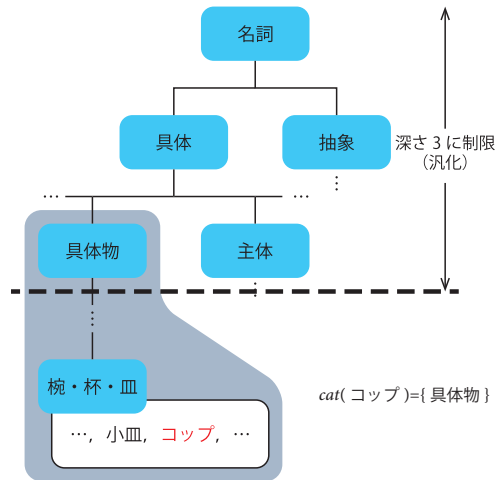


図 4.2 名詞意味属性

対象動詞の意味類型候補 意味類型辞書中で対象動詞に付与されている意味類型候補を素性とする。4 種類の意味類型について、対象動詞にその意味類型が候補として付与されていれば 1、そうでない場合には 0 の 2 値をとる。

$$\phi_{t_i}(s, t) = \begin{cases} 1 & \text{if } types(w_i) \ni t_i \\ 0 & \text{otherwise} \end{cases} \quad (4.3)$$

ただし、 $types(w)$  は動詞  $w$  に付与された意味類型候補、 $t_i \in T$  ( $1 \leq i \leq |T|$ ,  $|T| = 4$ ) であり、 $T$  は全意味類型を表す。

文中の格助詞および係助詞 日本語の格は格助詞または係助詞によって決定されるため、文中での格助詞・係助詞の出現を素性とする。IPA 辞書中に含まれる全 118 の格助詞・係助詞について、文中に含まれれば 1、そうでない場合には 0 の 2 値をとる。

$$\phi_{p_i}(s, t) = \begin{cases} 1 & \text{if } s \ni p_i \\ 0 & \text{otherwise} \end{cases} \quad (4.4)$$

ただし、 $p_i \in P$  ( $1 \leq i \leq |P|$ ,  $|P| = 118$ ) であり、 $P$  は IPA 辞書に含まれる全格助詞・係助詞を表す。

文中の名詞の意味属性の頻度 文中に出現する名詞の意味属性の頻度を素性とする。本研究では、名詞の意味属性情報として日本語語彙大系 [7] の名詞意味属性および固有名詞意味属性を用いた。日本語語彙大系の名詞・固有名詞意味属性は 3,000 属性からなるシソー

入力：  
六角堂の前にある、蕪村庵というお店に行ってきた。

出力：

六角	名詞,固有名詞,人名,姓,*,*	六角,ロツカク,ロツカク,*
堂	名詞,固有名詞,人名,名,*,*	堂,ドウ,ドー,*
の	助詞,連体化,*,*,*	の,ノ,ノ,*
前	名詞,副詞可能,*,*,*	前,マエ,マエ,*
に	助詞,格助詞,一般,*,*	に,ニ,ニ,*
ある	動詞,自立,*,*	五段・ラ行,基本形,ある,アル,アル,形容
、	記号,読点,*,*,*	,、,、,*
蕪村	名詞,固有名詞,人名,名,*,*	蕪村,ブソン,ブソン,*
庵	名詞,一般,*,*,*	庵,アン,アン,*
という	助詞,格助詞,連語,*,*	という,トイウ,トユウ,*
お	接頭詞,名詞接続,*,*	お,オ,オ,*
店	名詞,一般,*,*	店,ミセ,ミセ,*
に	助詞,格助詞,一般,*,*	に,ニ,ニ,*
行っ	動詞,自立,*,*	五段・カ行促音便,連用夕接続,行く,イツ,イツ,動作
て	助詞,接続助詞,*,*	て,テ,テ,*
き	動詞,非自立,*,*	カ変・クル,連用形,くる,キ,キ,*
た	助動詞,*,*	特殊・タ,基本形,た,タ,タ,*
。	記号,句点,*,*	。 ,。 ,。 ,*
EOS		

図 4.3 意味類型曖昧性解消器の入出力例

ラスになっており,計 30 万語についての意味属性の情報が付与されている。本研究ではデータスパースネスの問題を避けるため,意味属性の汎化を行いシソーラスの深さを 3 までに制限した(図 4.2)。この汎化により意味属性種類数は 21 種類に減少する。また,多義性を持つ名詞は複数の意味属性を持つことがあるため,頻度の計算の際には(ある名詞の文中の出現頻度) / (その名詞が持つ属性数)をその頻度として用いる。

$$\phi_{c_i}(s, t) = \sum_{n \in \{n_j | n_j \in N, \text{cat}(n_j) \ni c_i\}} \frac{|\{w_k | w_k \in s, w_k = n\}|}{|\text{cat}(n)|} \quad (4.5)$$

ただし,  $\text{cat}(n)$  は名詞  $n$  がもつ汎化した全意味属性,  $N$  は全ての名詞,  $c_i \in C$  ( $1 \leq i \leq |C|, |C| = 21$ ) であり,  $C$  は全ての汎化した意味属性である。

本研究では教師あり機械学習に LIBSVM[8] を用いた。この際のカーネル関数には線形カーネルを用い,分類精度を最大化するようなパラメータに設定した。

意味類型曖昧性解消器は,動詞の意味類型情報を形態素解析結果に付与する形で出力する。図 4.3 にその入出力例を示す。形態素解析器 MeCab のデフォルトのフォーマットの各行の末尾に,推定した意味類型情報の列を追加している。該当行が動詞でない場合には\*印が出力される。

表 4.1 評価データ中の意味類型の頻度

動作	変化	感覚・感情	形容	合計
451	614	255	680	2,000

表 4.2 意味類型曖昧性解消評価実験結果

手法	精度 [%]
SVM	69.9
random	24.1
majority	30.1

表 4.3 意味類型曖昧性解消器の評価実験結果内訳

正解	予測				計	再現率 [%]
	動作	変化	感覚・感情	形容		
動作	276	64	61	50	451	61.2
変化	31	435	10	138	614	70.8
感覚・感情	63	28	147	17	255	57.6
形容	23	99	18	540	680	79.4
計	393	626	236	745	2,000	-
適合率 [%]	70.2	69.5	62.3	72.5	-	69.9

## 4.2 評価・考察

作成した意味類型曖昧性解消器を評価するため、新聞記事を用いて評価実験を行う。2004年の日本経済新聞 [9] の中から複数の意味類型候補を持つ動詞を 1 つ含む文 2,000 文を無作為抽出し、人手で最も適切な意味類型を付与することで評価データを作成した。評価データ中の意味類型の頻度は表 4.1 の通りである。

作成した意味類型曖昧性解消器は、10 分割交差検定を用いて評価する。また、比較のために以下に示す 2 種類のベースラインを設定した。

**random** 対象動詞の意味類型候補から無作為に 1 つの意味類型を選択する

**majority** 対象動詞の意味類型候補から評価データ中で再頻出の意味類型を選択する

評価実験の結果を表 4.2 に示す。表中の手法における SVM は作成した意味類型曖昧性解消器、random、majority はそれぞれ該当するベースラインを表す。表より、作成した意味類型曖昧性解消器はベースラインの精度を上回り、約 70% の精度を達成したことが確認できる。表 3.1 より、全動詞の約 80% が単一の意味類型候補を持っていることを考慮すると、全動詞に対して 90% 以上の精度で正しい意味類型を付与することができるようになったと言える。また、意味類型曖昧性解消器の評価実験結果の内訳を表 4.3 に示す。表より、適合率と再現率の両方において意味類型の感覚・感情に対する性能が最も悪いことが確認できる。これは表 4.1 より、感覚・感情に対する学習データの量が少ないことが

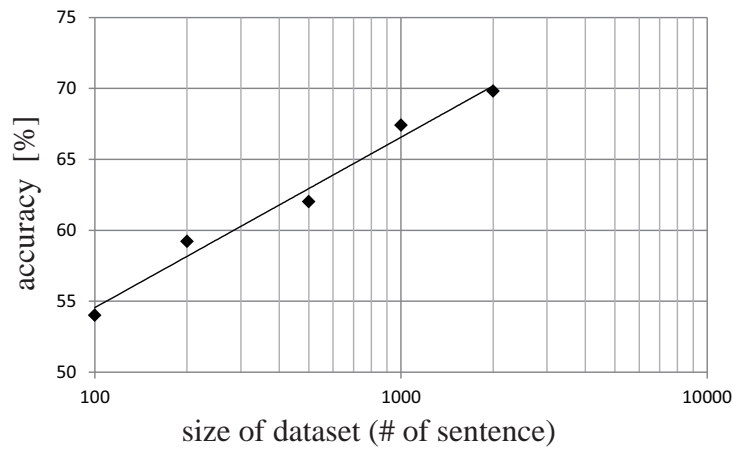


図 4.4 学習データ量の変化に対する意味類型曖昧性解消器の精度の変化

原因であると考えられる。このことから、感覚・感情に対する学習データを増やすか、SVMの学習の際に重み付け学習を用いることで意味類型曖昧性解消器の性能を向上させることができると思われる。

また、学習データの量を変化させた時の意味類型曖昧性解消器の精度の変化を図 4.4 に示す。図より、学習データが 2,000 文においても精度が上昇傾向にあるため、学習データの量を増加させることで、さらに精度向上が見込めると考える。

作成した意味類型曖昧性解消器は公開する。

## 5 サ変名詞に対する意味類型候補付与

意味類型辞書は一般動詞にしか意味類型候補が付与されておらず、サ変動詞に対しては付与されていない。これはサ変動詞が IPA 辞書の品詞体系上ではサ変名詞 + 「する」という形で扱われる（サ変動詞という品詞が存在しない）ためである。辞書中に意味類型候補が付与されていなくても、すべての意味類型を候補として意味類型曖昧性解消器を用いることで文中のサ変動詞に対して意味類型を 1 つに決定することも可能であるが、精度向上のためにできるだけ多くのサ変名詞に対して意味類型候補を付与することが望ましい。そこで本章では換言辞書を用いてサ変名詞に対して換言を行い、換言後の表現と意味類型辞書の情報に基づいて意味類型の付与を行う。

### 5.1 用言等換言辞書

用言等換言辞書は、サ変名詞、動詞、形容詞、副詞を人手で換言した辞書である。サ変名詞については、そのサ変名詞が「する」を伴った際（つまりサ変動詞）の意味で換言が行われる。例えば以下のような換言の情報が含まれている。

- 成長（する）      育つ
- 右折（する）      右に曲がる

本章では、換言後の表現に含まれる動詞の意味類型候補を付与するという基本的なアイデアに基づいてサ変名詞への意味類型候補の付与を行う。ただし、付与に際して次のような問題点が存在する。

- (1) 換言辞書と意味類型辞書で基にしている辞書が異なる
- (2) 換言辞書は換言なしを許容している
- (3) 換言後の表現は動詞 1 語とは限らない
- (4) 換言後の表現にサ変動詞が含まれる場合がある

1 つ目の問題点は登録されている見出し語に差異があることを意味している。これは 2 つ目の問題点と合わせて、全てのサ変名詞を換言することができないことを表している。3

つ目の問題点は、換言語の表現が最大2~3語からなる句になり得ることに起因しており、2つ以上の動詞を含む場合や、逆に動詞を含まない場合ある。4つ目の問題点は、換言後の表現がサ変動詞である場合には依然として意味類型候補が付与できないことを表している。

これらの問題に対応するため、次節に示す手順に従ってサ変名詞への意味類型候補の付与を行う。

## 5.2 サ変名詞への意味類型候補付与手順

サ変名詞への意味類型候補付与の流れを5.1に示す。まず、対象のサ変名詞をに対して換言辞書を用いてサ変動詞を含まない表現に換言する。ただし、換言先が再びサ変動詞を含む表現となった場合には、そのサ変動詞を再度換言する。再度換言後に次の条件に当てはまる場合は、元のサ変名詞の意味類型候補は人手により付与する。

- 換言先が再びサ変動詞になった場合
- 換言先の語が元のサ変名詞から意味が変わってしまっていると判断された場合

前者の場合に人出付与を行うのは2回以上換言すると、意味が大きく変わってしまうと考えるためである。また、換言辞書中に見出し語が存在しない、または、換言が存在しないサ変名詞については全ての意味類型候補を付与する。

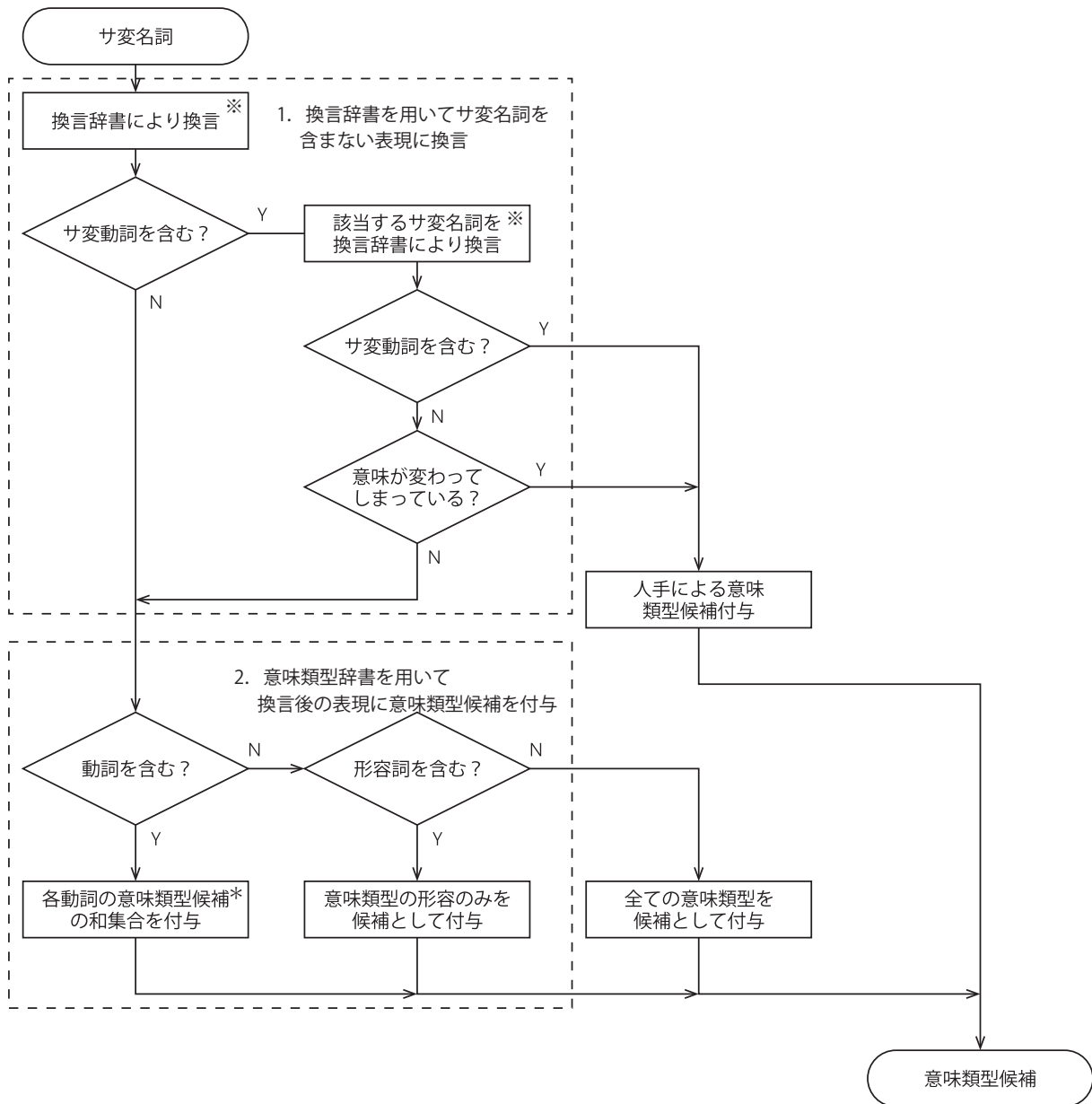
換言を行うことができた場合、換言後の表現に対して以下に示す規則に従い意味類型を付与する。

### 5.2.1 意味類型候補付与規則

換言後の表現に対して MeCab を用いて形態素解析を行い、その結果に対して上から順番に当てはまるものを適用する。

- (1) IPA 辞書の品詞体系において「動詞, 自立」の語を1つ以上含む場合、該当するすべての動詞の意味類型の和集合を付与
- (2) 品詞が「形容詞」の語を1つ以上含む場合、「形容」の意味類型のみを付与
- (3) 「動作」「変化」「感覚・感情」「形容」のすべての意味類型を付与

ただし、換言先に1つでも中山辞書中にない動詞を含んでいた場合は人手で意味類型を付与しなした。換言先に複数の動詞を含む際に意味類型の和集合を取る理由は、複数の意味類型から1つに特定するのは意味類型曖昧性解消器の役割であり、候補付与の時点ではとり得る意味類型を可能な限り列挙することが目的であるからである。



※換言が存在しない場合は全ての意味類型を候補として付与する。  
 \*対象動詞が意味類型辞書に含まれていない場合は、サ変名詞に対して人手で意味類型候補の付与を行う。

図 5.1 サ変名詞への意味類型候補付与の流れ

また、1つのサ変名詞が複数の換言先を持つ場合はそれぞれの換言先に対する意味類型の和集合を最終的な意味類型とする。

具体的な意味類型候補付与例を以下に示す。

例 1 おしゃべり(する)      話す：動作

1度の換言で意味類型候補付与に成功している



表 5.1 サ変名詞 3,934 語に対する意味類型候補種類数の内訳

種類数	語数
1 種類	2,872 語
2 種類	718 語
3 種類	295 語
4 種類	49 語

例 2 ジャッジ(する) 判断する 決める：感覚・感情

再換言により意味類型候補の付与に成功している

例 3 てんてこ舞い(する) 形容：形容

換言先が形容詞であるため，形容のみを付与している

例 4 アピール(する) 主張する 言う：×

再換言後，意味が変わってしまっていると判断（人出による付与）

例 5 執刀(する) 手術する 治療する：×

再換言してもサ変動詞のまま（人手による付与）

## 5.3 結果

IPA 辞書中のサ変名詞 12,041 語に対して，換言辞書中に換言が存在したのは 3,934 語であった．その 3,934 語に対して先述の手順に従って意味類型候補付与した結果，意味類型候補種類数の内訳は表 5.1 の通りになった．この表より全サ変名詞の 1/3 に当たる約 3,000 語については 1 種類の意味類型のみが付与され，曖昧性解消の必要が無くなったことが確認できる．また，約 1,000 語についても意味類型の候補が 2 種類ないし 3 種類になったことで，全ての意味類型を候補とする場合と比べて曖昧性解消の精度向上に寄与していると考えられる．

また，意味類型候補をどれだけ正しく付与できたかの評価を行った．付与を行ったサ変名詞 3,934 語から無作為に抽出した 100 語に対して著者 1 人の主観により意味類型候補を手動付与し<sup>1</sup>，換言辞書を利用して自動付与された意味類型候補と比較を行った．その結果を表 5.2 に示す．表より，自動付与したサ変名詞のうち約 8 割については少なくとも必要な意味類型候補が付与されていることが分かる．自動付与であるにもかかわらず 8 割程度の精度で意味類型候補を付与できているのは，人手により構築されたノイズの少ない

<sup>1</sup>この際に，サ変名詞の意味がわからない場合に辞書等を参照してよいとした．

表 5.2 換言辞書を用いた自動付与と手動付与の意味類型候補の比較

	サ変名詞数 (100 語中)
手動付与と自動付与の結果が完全一致	64
自動付与の結果が手動付与の結果を内包	82
手動付与と自動付与の結果に重なりがない	8

表 5.3 手動付与と自動付与の結果に重なりがなかったサ変名詞の内訳

サ変名詞	換言後の表現	手動付与結果	自動付与結果
受容	<u>受け入れる</u>	動作	変化, 感覚・感情
生息	<u>生きている</u>	変化	形容
荒廃	<u>荒れる</u>	動作	変化
起工	<u>工事を始める</u>	動作	変化
超越	<u>越える</u>	動作	変化, 形容
迫害	<u>苦しめる</u>	感覚・感情	動作
逆行	<u>戻る</u>	変化	動作
黙殺	<u>黙っている</u>	動作	感覚・感情

言語資源を、想定された利用方法の範囲内で利用しているためである。手動付与と自動付与の結果に重なりがなかった 8 個のサ変名詞を表 5.3 に示す。いずれの例も換言は妥当であり、対応する意味類型辞書のエントリを更新することにより解決すると考える。特に、動詞「生きる」に変化の意味類型候補が付与されているのは、作業時の誤りの可能性が高い。

サ変名詞に対して付与した意味類型候補は、手動付与した 100 件の内容を反映した上で公開する。

## 6 評判分析への応用

本章では意味類型の応用例として評判文抽出を扱う。評判表現の多くは「良い」「～が多い」などの形容表現が用いられる。意味類型の情報をを用いることで「～に優れる」といった動詞の形容表現を扱うことができる、また「～を感じる」「～だと思う」のような意味類型の感覚・感情に対応する表現も評判になりやすい。このようなことから、意味類型情報を利用することでより網羅性高く評判表現を扱うことができ、性能向上に貢献できるという仮説を立てた。

### 6.1 使用したデータ

対象とするコーパスには 249 記事、4186 文からなる人手による解析済みブログコーパスである KNB コーパス [4] を用いた。本研究における評判表現の定義は KNB コーパスに従い「ある対象に対して述べられた肯定的、もしくは否定的判断や態度、叙述」とする。KNB コーパスは評判表現についてアノテーションされており、「感情」「批評」など 7 種の評判タイプや評価の極性、評価保持者、評価対象といった詳細な情報も付与されている。しかし、本研究の目的は純粋な品詞情報と意味類型情報の比較であるため、評判表現の詳細な情報についての予測・同定は行わない。本研究では KNB コーパス中で評判表現を含む文を評判文、含まない文を非評判文と定義し、各文を評判文/非評判文に分類する 2 値分類タスクを考える。

### 6.2 評価手法

品詞情報のみを用いた場合と意味類型情報を用いた場合とを比較するため以下に示す手法を用いて実験を行う。

まず、入力文を MeCab を用いて形態素解析したあと以下に示す形態素解析の後処理を行う<sup>1</sup>。

<sup>1</sup>KNB コーパスは形態素アノテーション済みであるが、意味類型曖昧性解消器の入力が IPA 品詞体系日本語辞書の品詞体系に基づいていることを前提としているため再度形態素解析を行っている。

- 直後に原形が「する」の動詞を伴うサ変接続の名詞は動詞として扱う
- 直後に助詞「を」と原形が「する」の動詞を続けて伴うサ変接続の名詞は動詞として扱う
  - － 上記 2 つに当てはまった場合、動詞「する」は用言として扱わない
- 直後に原形が「だ」「です」の助動詞を伴う形容動詞語幹の名詞は形容詞として扱う
- IPA 品詞体系日本語辞書上の品詞細分類が自立でない動詞（「～している」の「いる」や「～し得る」の「得る」など）は用言として扱わない

この後処理はそれぞれ次に示す目的に基づく<sup>2</sup>。

- 意味類型曖昧性解消を「する」ではなく、語義を担うサ変名詞に対して行う
- 形容動詞（形容動詞語幹（名詞）＋「だ/です」（助動詞））を形容詞として扱う
- 事実上機能語として振る舞う動詞を用言から除外する

その後、用言として扱われる動詞に対しては意味類型曖昧性解消器を用いて意味類型を付与し、形容詞に対しては意味類型の「形容」を常に付与する。

次に品詞情報、または意味類型情報を用いてコーパス中の全文から評判文の抽出を行いそれぞれの結果を比較する。先述のとおり評判表現の多くは形容表現が用いられていること、また、動作を表す表現は結果として客観的な事象を表していると考えられることから、以下に示す規則を用いて評判文の抽出を行う。

**品詞情報のみを用いた規則** 文中の最後に出現する用言の品詞が動詞であれば非評判文であると、形容詞であれば評判文とする。

**意味類型情報を用いた規則** 文中の最後に出現する用言の意味類型が動作であれば非評判文であると、変化、感覚・感情、形容のいずれかであれば評判文とする<sup>3</sup>。

文中の 1 つの用言だけを見るという単純な規則を用いるのは、結果に寄与する要因を減らし品詞情報と比較した意味類型情報の特徴を明らかにする意図に基づいている。また、最後の用言に注目しているのは、日本語文の主要部は後ろに出現しやすいという仮定に基づいている。

<sup>2</sup>なお、この後処理を行わずに品詞情報のみを用いて評判文抽出を行う場合、後述する結果と比較して悪い結果を示した。具体的には分類精度が 0.574、F 値が 0.310 となった。

<sup>3</sup>意味類型が変化、感覚・感情の場合に評判文とするのは予備実験の結果に基づく。

表 6.1 品詞情報のみを用いた評判文抽出

正解\予測	評判文	非評判文	再現率 [%]
評判文	533	1512	26.1
非評判文	177	1964	91.7
適合率 [%]	75.1	56.5	59.7

評判文抽出の観点からの F 値=38.7%

表 6.2 意味類型情報を用いた評判文抽出

正解\予測	評判文	非評判文	再現率 [%]
評判文	1442	603	70.5
非評判文	889	1252	58.5
適合率 [%]	61.9	67.5	64.4

評判文抽出の観点からの F 値=65.9%

なお { 品詞が動詞の用言 } ⊃ { 意味類型が動作の用言 } であるので，意味類型を用いた場合は必ず評判文に対する再現率が上昇することに留意されたい。

## 6.3 結果・考察

表 6.1, 表 6.2 に品詞情報のみを用いた場合と意味類型情報を用いた場合の評判文抽出結果をそれぞれ示す。意味類型情報を用いた場合，品詞情報のみを用いた場合と比べて評判文/非評判文の分類精度において 4.7 ポイント，評判文抽出の観点から見た F 値において 27.2 ポイント上昇している。

### 6.3.1 誤り解析

品詞情報のみでは誤りで，意味類型情報を扱うことで正しく評判文であると判断された例は 909 文存在した。誤りには「～がわからなかつたです。」「～と思います」「～と感じました」など，意味類型が感覚・感情の動詞で終わる文を多く含んでおり，動詞をその意味によって分類するという意味類型導入の意図が反映されており，有用性が確認できた。

品詞情報，意味類型情報どちらを用いた場合にも共通して誤りだった例は 780 文存在した。うち評判文を非評判文としてしまった例は 603 文，非評判文を評判文と判定してしまつた例は 177 文だった。さらに前者について誤りの原因の解析を行った。全 603 文から 50 文を無作為に抽出した解析を行った結果の内訳は以下のとおりである。

意味類型が動作でも評判表現になる例 (16 件) 以下に示す表現においては下線部の動詞<sup>4</sup>の意味類型によらず評判表現になり得る。このような誤りはパターンマッチで個別に扱うことで解決できる。

「～してしまう」「～れた」「～してくれる」「～しか～しない」「～してみ

<sup>4</sup>下線部が 1 動詞を表す。「～する」表記だがサ変動詞に限らない。

たい」「～しやがって」「～しないでくれ」「～しない」「～すべきではないだろうか」「～して行きたい」

また、「普段から使っている」「観光客が押し寄せる」「騒ぎ散らす」のように特定の用法で評判表現になる例も存在した。

用言以外が評判表現（のトリガ）になる例（12件） 用言以外の語にも評判表現になる、あるいはその語を伴うことではじめて評判表現になる、という語が存在する。具体的には「それ以上のもの」「宝もの」「人だかり」「バカ」「抜群」といった名詞、「ぜひ」「あっさり」「日常的に」といった副詞、「ありがとう」「ジーザス」といった感動詞、「超」といった接頭辞が確認された。「抜群」や「バカ」といった名詞に関しては形容動詞語幹として辞書に登録された方が良いと思われる。

評判表現が文の途中に出現する例（11件） 1文に複数の節が含まれているなどの理由により、文の途中の用言が評判表現になっている場合がある。例えば「さらに冬場は肌が荒れやすいので、レバーの生姜煮を追加する。」という文において「荒れやすい」は評判表現であるが、文末のサ変動詞「追加する」が意味類型の動作を持つため本実験で用いた手法では非評判文と判断された。このような事例は係り受け解析や述語項構造解析を行い、その情報を用いることで扱うことができる。

形態素解析誤りに起因する誤り例（7件） うち4件はブログ記事の文体に起因するものであった。その具体例を以下に示す。

例1 いくら特売だったからといって3斤は身がもたねー!

「もたない」の話し言葉である「もたねー」が正しく解析されなかった。

例2 種類は豊富だし、なんと言ってもあっさり系が多い!!!

形容詞「多い」の「い」が小文字化されたことで正しく解析されなかった。

例3 食べ物の魅力はオソロシイ。(カタカナ化)

形容詞「恐ろしい」がカタカナ化されたことで正しく解析されなかった。

例4 最初はカメラもしょぼかったし、画面も小さかったし、音も3和音とかやった。

「だった」の方言である「やった」が動詞（「やる」の連用形）として解析された。

形容動詞語幹の扱いに起因する誤り例（3件） 直後に「だ」や「です」を伴わない場合でも、形容表現として形容動詞語幹が用いられている場合がある。「～がお洒落。」「～が豊富。」「～って素敵。」のように文末に出現する場合は該当する。

表 6.3 KNB コーパス中の評判表現に含まれていた動詞「ある」の項

意味\項	名詞単体	名詞句
(属性を)有する	楽しみ, 問題, 雰囲気, 制限, 弱点, 開き, 魅力, おそれ, 価値, 歴史, 見所, ご利益, 種類, 威厳, 義務, 由来, メリット, 人気, チャンス, オチ, 距離, 使い道, 違和感, 文化, 割引, 安心感	例) 見直す必要, 得るもの, QOL を高める要素, 色々な機能, 参考になる点, 歴史を感じさせるもの, 興味深いもの, 引き込まれるもの
存在する	自然, 名所, 見所	例) いいところ, 有用なもの, 素敵なカフェ, お勧めスポット
(事象)が起こる	火事	例) 迷うこと, 無料になる場合, 不安になるとき
感情	興味, 文句, 思い入れ, 不満, 意見	例) そんなに使わなかった記憶, 面白いことになりそうという意識, 納得した覚え, 多いという印象

意味類型付与誤りに起因する例(1件) 意味類型を付与する際に誤った意味類型が付与されてしまい, 結果として誤った例が1件存在した。

品詞情報のみでは正しかったが意味類型情報を扱うことによって本来は非評判文である文を評判文と判断し誤りとなった例は712文存在した。その中で最も大きい割合(2割程度)を占めていたのが, 文中の最後の用言が動詞「ある」の場合である。動詞「ある」は存在を表すので意味類型は常に形容であるが, 実際に評判表現になるか否かは項の名詞(句)に依存する。

評判文になる例

レトロで綺麗なお庭があります。

非評判文になる例

~, その先に町家風の店内がある

表6.3にKNB コーパス中の評判表現に含まれていた動詞「ある」の項を示す。表中で「例」と書かれていない項目は, コーパス中で見つかった全ての項を挙げている。項が名詞単体の場合, 評判表現になるか否かは名詞自身の持つ極性によって決まると考える。「~がある」を伴った際の名詞の評価極性情報は日本語評価極性辞書(名詞編)<sup>5</sup>に含まれており, これを利用することで評判表現か否かを判別することが可能になると考えられる。実際に表6.3中の名詞単体のうち「由来」「使い道」「オチ」「割引」「意見」を除いた全ての名詞が日本語評価極性辞書(名詞編)に含まれていた。項が名詞句の場合は主辞である名詞に

依存する場合と、修飾部に依存する場合に分けられる。項の主辞が「機能」「印象」などの特定の名詞の場合は修飾部に依らず評判表現になりやすいが、「もの」や「ところ」あるいは「カフェ」のような単体では評判に結びつきにくい名詞が主辞の場合は評判表現か否かが修飾部に依存すると考えられる。

### 6.3.2 変化の意味類型が評判分析に与える影響

本実験では文中の最後の用言の意味類型が変化の場合は評判文であると判定して、そのような規則に設定したのは予備実験の結果であるが、実際に意味類型が変化である用言がなぜ評判表現になり得るかを分析した。

まず、文中の最後に出現する、意味類型が変化の動詞の内訳を調べた。その結果、動詞「なる」が最も大きい割合を占めており 485 件中の約 1/5 にあたる 106 件が動詞「なる」であった。さらにそのうち評判文であったのは約 6 割の 64 件であった。このことから、意味類型が変化であるという情報だけでは評判表現かどうかを判断することは難しく、また、その中で動詞「ある」がどのような表現中で使われると評判表現になるのかという情報が重要であると考えられる。

動詞「なる」を含む表現が評判表現になる場合について観察するとその多くは表層格二格を持つ場合と、用言の連用形と共起する場合に大別できることがわかった。表層格二格については、例えば「いい思い出になる」のような表現は評判表現になり得る。どのような名詞(句)が二格になった場合に評判表現になり得るかについては、前節の動詞「ある」の主辞と同様の議論が成立すると考える。用言の連用形と共起する場合には「煮込む系の料理がしたくなる」「心も暖かくなります」のような場合が評判表現になり得る。評判表現になるかどうかを判定するには、用言の種類やパターンマッチが手がかりになると考える。例えば、形容詞連用形 + 「なる」(意味類型：変化) というパターンは評判表現になりやすいと考える。



## 7 結論

本研究ではまず、意味類型の実応用を目的として動詞意味類型曖昧性解消器の作成を行った。意味類家曖昧性解消器とは文中の動詞の意味類型を1つに決定するものであり、形態素解析の結果と既存の言語資源から得られる情報を素性とした教師あり機械学習アプローチを用いて作成した。その結果、複数の意味類型候補を持つ動詞に対して70%の精度で意味類型を決定することができ、全動詞に対しては90%以上の精度で正しく1つの意味類型を決定することができるようになった。

次に、意味類型の概念をサ変動詞にも適用することを考え、サ変名詞への意味類型候補の付与を行った。用言等換言辞書を用いてサ変名詞を換言し、換言後の表現に含まれる用言を手がかりにすることでサ変名詞へ意味類型候補を付与した。その結果、全サ変名詞の約1/3にあたる3,934語に対して意味類型候補を付与することができた。

最後にこれまでの研究成果を用いて意味類型を評判分析に応用し、その有用性を確認した。評判表現が含まれている文を評判文、含まれていない文を非評判文としブログ記事中に含まれる各文に対して、評判文/非評判文の2値分類を行う評判文抽出タスクを設定し、ルールベース手法において品詞情報を用いた場合と意味類型情報を用いた場合とで比較を行った。その結果意味類型情報を用いた場合、品詞情報を用いた場合と比較して分類精度で4.7ポイント、評判文抽出のF値で27.2ポイント上昇し、意味類型の有用性を確認することができた。

意味類型はこれまで動詞・形容詞という品詞情報を用いていた自然言語処理タスクにおいて品詞情報の代わりに利用することで、語の意味をより正確に扱うことができるようになるため、その波及効果は大きい。これらの研究成果は言語資源・ツールの形で公開され使用可能になる。

## 謝辞

本研究を遂行するにあたり，多大なる御指導そして御協力いただきました長岡技術科学大学の山本和英准教授に深く感謝致します．山崎克之教授ならびに王龍標産学融合特任准教授には，論文の推敲に際し貴重な御意見をいただきました．深く感謝致します．

最後に，様々な場面でお世話になりました山本研究室の皆様に，心より感謝致します．

## 使用したツール及び言語資源

- [1] IPA 品詞体系日本語辞書 . Ver. 2.7.0 . <http://mecab.sourceforge.net/src/>
- [2] 中山匠, 山本和英 . SNOW D1: 動詞意味類型辞書 . <http://www.jnlp.org/SNOW/D1>
- [3] 吉倉孝太郎, 山本和英 . SNOW D2: 内容語換言辞書 . <http://www.jnlp.org/SNOW/D2>
- [4] 橋本力, 黒橋禎夫, 河原大輔, 新里圭司, 永田昌明 . 構文・照応・評判情報つきプロ  
グコーパスの構築 . 自然言語処理 Volume 18, Number 2, pp. 175-201 . 2011 .
- [5] 東山昌彦, 乾健太郎, 松本裕治 . 述語の選択選好性に着目した名詞評価極性の獲得 .  
言語処理学会第 14 回年次大会論文集, pp. 584-587 . 2008 .
- [6] 工藤拓 . 形態素解析器 MeCab. Ver. 0.996 . <http://mecab.sourceforge.net/>
- [7] 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 CD-ROM 版 . 岩波書店 . 1999 .
- [8] Chih-Chung Chang, Chih-Jen Lin. LIBSVM. Ver. 3.18. [http://www.csie.ntu.edu.  
tw/~cjlin/libsvm/](http://www.csie.ntu.edu.tw/~cjlin/libsvm/)
- [9] 日本経済新聞全記事データベース CD-ROM 版, 2004 年版 .

## 参考文献

- 1) Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet Project. In *Proceedings of the 1998 Joint Conference on 36th ACL and 17th COLING*, pp. 86–90, 1998.
- 2) Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of WWW 1998*, pp. 107–117, 1998.
- 3) Danilo Croce, Alessandro Moschitti, Roberto Basili, and Martha Palmer. Verb Classification using Distributional Similarity in Syntactic and Semantic Structures. In *Proceedings of the 50th ACL*, pp. 263–272, 2012.
- 4) Steve L. Manion and Raazesh Sainudiin. An iterative ‘sudoku style’ approach to subgraph-based word sense disambiguation. In *Proceedings of \*SEM 2014*, pp. 40–50, Dublin, Ireland, August 2014. Association for Computational Linguistics and Dublin City University.
- 5) Tetsuji Nakagawa, Kentaro Inui, and Sadao Kurohashi. Dependency Tree-based Sentiment Classification using CRFs with Hidden Variables. In *Human Language Technologies: NAACL 2010*, pp. 786–794, 2010.
- 6) Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd ACL*, pp. 115–124, 2005.
- 7) Thomas Scholz and Stefan Conrad. Opinion Mining in Newspaper Articles by Entropy-Based Word Connections. In *Proceedings of the EMNLP 2013*, pp. 1828–1839, 2013.
- 8) Karin Kipper Schuler. VerbNet: A broad-coverage, comprehensive verb lexicon, 2005.

- 9) Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In *Proceedings of the EMNLP 2013*, pp. 1631–1642, 2013.
- 10) Hiroya Takamura, Takashi Inui, and Manabu Okumura. Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd ACL*, pp. 133–140, 2005.
- 11) Xiaodan Zhu, Hongyu Guo, Saif Mohammad, and Svetlana Kiritchenko. An Empirical Study on the Effect of Negation Words on Sentiment. In *Proceedings of the 52nd ACL*, pp. 304–313, 2014.
- 12) 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. *自然言語処理*, Vol. 12, No. 3, pp. 203–222, 2005.
- 13) 倉島健, 藤村考, 奥田英範. 大規模テキストからの経験マイニング. *電子情報通信学会論文誌*, Vol. 92, No. 3, pp. 301–310, mar 2009.
- 14) 竹内孔一, 土山傑, 守屋将人, 森安祐樹. 類似した動作や状況を検索するための意味役割及び動詞語義付与システムの構築. *電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション*, Vol. 109, No. 390, pp. 1–6, 2010.
- 15) 谷口秀治. 日本語教育における動詞の分類について. *大分大学留学生センター紀要 2号*, pp. 53–63, 2005.
- 16) 金田一春彦. 日本語動詞の一分類. *日本語動詞のアспект*, pp. 5–26. むぎ書房, 1976 (原論文 1950).
- 17) 中山匠, 山本和英. 用言の新しい意味類型 —作用性用言と形状性用言—. *言語処理学会 第17回年次大会 発表論文集*, pp. 560–563, 2011.
- 18) 山本和英, 吉倉孝太郎. 用言等換言辞書を人手で作りました. *言語処理学会 第19回年次大会 発表論文集*, pp. 276–279, 2013.

## 本論文に関する原著論文

- 1] 岡田 正平，山本 和英．動詞意味類型の曖昧性解消に向けた格フレーム情報との関連調査．言語処理学会第 20 回年次大会，pp. 524-527，2014．
- 2] Shohei Okada and Kazuhide Yamamoto. Semantic Type Disambiguation for Japanese Verbs. In *Proceedings of the International Conference on Asian Language Processing 2014 (IALP 2014)*, pp. 6-9, 2014.
- 3] 岡田 正平，山本 和英．評判分析における品詞情報と意味類型情報の有効性比較．言語処理学会第 21 回年次大会，no page numbers，2015（発表予定）

## 付録 A : サ変名詞に付与した意味類型候補

換言辞書を用いたサ変名詞に対する意味類型候補の自動付与結果を、付与された意味類型候補ごとに掲載する<sup>1</sup>。ただし、動作のみが付与されたサ変名詞は全部で 1,985 件と数が多いため、無作為抽出した 500 件のみを掲載する。

### 動作（500 件無作為抽出）

お出かけ, キャッチ, コピー, コーチ, サイン, サポート, サーフィン, シミュレーション, スタート, ストップ, スピーチ, スライス, チャレンジ, チャージ, ディベート, デザイン, デート, トライ, トレーニング, ドライブ, ドロップ, ファックス, ミックス, ミーティング, ランニング, 一新, 一礼, 万引き, 上乘せ, 上塗り, 上奏, 下書き, 下調べ, 中断, 主催, 乗車, 乱伐, 交付, 交差, 交易, 介抱, 介護, 代用, 仲立ち, 休会, 休学, 休憩, 休診, 休養, 会得, 会葬, 会計, 伝聞, 作成, 作曲, 作製, 作詞, 使役, 供託, 供述, 促成, 保守, 保留, 修正, 停止, 催促, 償却, 先制, 先攻, 全休, 公称, 共感, 共生, 再審, 再選, 出張, 出発, 出社, 出荷, 出頭, 分与, 分担, 分類, 前借り, 前売り, 前記, 前進, 創出, 創建, 剽窃, 加入, 加護, 加速, 励行, 勘当, 勝負, 包囲, 北上, 区分, 協同, 南中, 博覧, 占領, 即応, 厳守, 参会, 反復, 反芻, 収穫, 収賄, 収集, 受刑, 受容, 受理, 受給, 受診, 受賞, 受験, 口出し, 口外, 口論, 口述, 召還, 合作, 合図, 合奏, 合戦, 合掌, 同期, 告知, 告示, 周旋, 味方, 呼応, 命令, 啓示, 喫煙, 嘲笑, 噂, 回遊, 図示, 圧延, 執筆, 報償, 増大, 売り買い, 外遊, 外食, 多作, 大拳, 大笑い, 大食, 失敬, 失火, 奇襲, 奏楽, 奔走, 奪還, 始末, 守備, 守護, 安住, 完納, 実行, 実証, 宿泊, 寄進, 密売, 密漁, 密猟, 密談, 寝泊まり, 審問, 審議, 対戦, 対話, 小売り, 巡視, 工作, 帰休, 帰国, 帰港, 常食, 幫助, 広言, 延命, 延期, 建国, 建議, 建造, 弁済, 強奪, 強弁, 強攻, 強襲, 形容, 待機, 徐行, 御参り, 御膳立て, 徴兵, 徴収, 徴集, 心労, 志向, 応接, 息抜き, 意識, 愛飲, 感嘆, 手分け, 手招き, 手書き, 扮装, 扶養, 承諾, 抑制, 抑止, 投与, 投薬, 折半, 拔擢, 抹殺, 拘束, 招聘, 招請, 拝聴, 持ち逃げ, 指向,

<sup>1</sup>論文中に記載した手法の結果であり、実際に公開される情報とは異なることに留意されたい。例えば、自動付与の評価のために行った手動付与の結果はこの付録に反映されていない。

挑戰，拳式，振動，振込，捕獲，搜查，採取，接地，推進，揭示，提携，提示，援助，搬送，摩耗，擊退，撤退，擁護，操業，支援，支給，改修，改善，改正，改称，放出，放水，放置，救助，救護，敢行，敢闘，散布，散見，整理，整頓，敷設，齊唱，施工，施策，早退，昇降，明示，明記，映写，暗算，暗躍，束縛，来航，来襲，查察，植樹，檢問，檢查，檢索，檢視，檢閲，模倣，横転，樹立，欠勤，欠航，歡談，止血，死闘，殴打，殺菌，水泳，洗淨，洗面，流浪，流用，消去，消火，添乗，減俸，減反，減点，湯治，滅菌，滑降，漂流，演技，漸進，潜入，無心，烧却，照会，爆破，牛飲馬食，特集，牽制，猪突，率先生食，申告，略取，痙攣，発売，発掘，発明，発表，発送，盗作，目撃，直談判，着信，着工，着服，着水，砲撃，破碎，秘蔵，移管，移築，移行，移送，種別，稽古，穴埋め，究明，空費，突破，突貫，突進，節約，築港，精製，組成，経営，絶賛，続行，編纂，縮小，翻訳，考查，聴取，育児，育成，脚色，脱穀，脱臭，自称，自給，自警，自賛，自足，船出，色刷り，苦闘，菜食，落書き，著作，行き来，表彰，表示，裁縫，装備，補償，補充，補導，製氷，製油，複写，見学，見聞，観光，観劇，観察，解放，解明，解消，解熱，解答，解題，触発，言明，計数，計測，計量，訓練，記帳，記念，記録，診察，証明，試食，試飲，誤用，誤解，誤認，調理，談笑，談話，論述，講義，講釈，識別，販売，貫徹，贊助，賛辞，贈与，贗作，越冬，跳躍，転回，転籍，転職，転記，転載，近道，返却，返本，返還，追加，追尾，追跡，退散，通算，通院，速報，造営，造成，連取，連戦，逮捕，進呈，進学，進展，進撃，進軍，遂行，運休，適用，遭遇，遮蔽，遵守，配付，配属，配布，配管，量産，量販，錯覚，鎮静，門前払い，閉館，開墾，開封，開拓，開放，開花，開通，開門，闊歩，防水，防音，陳述，隱蔽，集配，集録，静養，非難，順延，頒布，飛び火，飲食，養生，驅使，驀進

## 变化

カムバック，キス，ゴールイン，ダイエット，ダウン，バック，ユーターン，リコール，リストラ，リタイア，一番乗り，下船，下落，下車，下降，中退，乱発，乾燥，交代，交錯，他界，付属，代行，低下，低落，低迷，併合，併発，依存，倍增，倒壊，値下がり，停泊，停電，偶発，傾斜，充血，全壊，全快，全治，全滅，共栄，具現，円熟，再任，再婚，再発，再興，再起，凝固，凝結，凱旋，出没，出現，出生，切腹，到着，割愛，勃発，化合，化膿，半壊，卒倒，即死，厚着，参政，参集，及第，収斂，収縮，収録，合体，合併，合同，合格，合流，否決，喪失，噴出，噴射，四散，回歸，固辞，圧死，在留，埋没，埋蔵，堆積，増加，増幅，増水，増額，壊死，壊滅，変死，多発，大儲け，大破，夭折，失格，失業，失点，失神，失職，婚姻，婚約，存命，孵化，完了，完成，完結，完走，完



遂, 密生, 密集, 対峙, 小破, 山崩れ, 山積, 崩壊, 崩御, 崩落, 帰任, 帰着, 帰還, 干拓, 平癒, 廃業, 廃絶, 延着, 弛緩, 形成, 往生, 後戻り, 後略, 後退, 徘徊, 復位, 復元, 復員, 復学, 復帰, 復旧, 復職, 復興, 復調, 微増, 心中, 快復, 急増, 急死, 急減, 急落, 急逝, 懸垂, 成熟, 成育, 成長, 戦死, 戦没, 投降, 拒否, 拒絶, 捻挫, 接続, 摩擦, 撮影, 散乱, 散在, 散発, 早世, 早起き, 明滅, 普及, 暴落, 更迭, 枯渴, 根治, 梗塞, 歓待, 歓迎, 死亡, 死別, 死去, 死没, 死滅, 死産, 殉教, 殉職, 残存, 残業, 残留, 気絶, 水死, 永眠, 汚染, 汚濁, 決壊, 派生, 流行, 浸透, 消失, 消滅, 液化, 混入, 減収, 減少, 減数, 減水, 減退, 満了, 溶解, 滅亡, 滅失, 激増, 激減, 火傷, 点滅, 焼失, 熟成, 爆死, 犬死に, 獄死, 現出, 現存, 生存, 生息, 生育, 生誕, 生起, 生長, 産卵, 病死, 病没, 発光, 発現, 発生, 発祥, 発育, 発達, 目減り, 省略, 着地, 着岸, 着陸, 破損, 破断, 破滅, 破門, 硬直, 空回り, 突発, 窒息, 紅葉, 終了, 終息, 終業, 終演, 終結, 経過, 結成, 結集, 絶句, 絶命, 緩和, 繁栄, 繁殖, 繁盛, 繁茂, 罷免, 群生, 群発, 老衰, 腐乱, 腐敗, 膨張, 自刃, 自害, 自殺, 自決, 自滅, 自生, 若死に, 落成, 落着, 蒸発, 融解, 行商, 衰亡, 衰弱, 被災, 覚醒, 解任, 解職, 解雇, 討ち死に, 誕生, 調停, 負傷, 起床, 超過, 転倒, 転生, 轢死, 辞任, 辞職, 辞退, 退任, 退団, 退学, 退官, 退廃, 退役, 退行, 退陣, 逆戻り, 逆行, 途絶, 逗留, 逝去, 速成, 連発, 遅刻, 遅延, 遅滞, 達成, 野合, 野垂れ死に, 金儲け, 錯綜, 長持ち, 長生き, 閉会, 閉幕, 閉廷, 降下, 降伏, 降板, 降格, 降着, 降車, 除名, 除隊, 隔絶, 集合, 集結, 雇用, 離婚, 離職, 雨漏り, 静止, 順応, 頻発, 顕在, 飛散, 飢え死に, 餓死, 黄葉

## 感覚・感情

あたふた, いかさま, うじうじ, うずうず, うっかり, おわび, イメージ, エンジョイ, ジャッジ, ネーミング, ビックリ, リクエスト, リラックス, 一喝, 一考, 予定, 予想, 予感, 予断, 予期, 予測, 予知, 予習, 予見, 交歓, 仮定, 仮想, 仰天, 企図, 企画, 促進, 保障, 信仰, 信奉, 信用, 信賴, 倦怠, 優勝, 先勝, 先決, 克服, 全勝, 全敗, 八百長, 公定, 公約, 公認, 共謀, 内省, 再考, 凋落, 処罰, 分析, 判別, 判定, 判断, 判明, 制定, 制限, 力負け, 加味, 努力, 動揺, 動転, 勘弁, 勘案, 勝利, 勝訴, 危惧, 即断, 即決, 厳戒, 可決, 叱責, 右往左往, 右顧左眄, 合算, 合計, 同情, 同感, 吟味, 否認, 呻吟, 呼称, 命名, 命拾い, 哀悼, 哀惜, 善戦, 嘆願, 四苦八苦, 回想, 回顧, 困惑, 困憊, 固執, 土下座, 圧勝, 執心, 執着, 堪忍, 堪能, 夢想, 大勝, 大敗, 奮戦, 威嚇, 威圧, 嫉妬, 嫌悪, 安堵, 安息, 完勝, 完敗, 定義, 実感, 客観, 宣誓, 容認, 察知, 審判, 対抗, 対比, 専従, 専心, 専念, 尊重, 尻込み, 屈従, 屈服, 屈託, 崇拜, 工面, 帰属, 帰

結，歸順，平伏，平身低頭，幻惑，幻想，度外視，強迫，当惑，後悔，從事，從屬，心服，心痛，忌避，忍從，忍耐，忘却，快勝，念願，怒号，思案，思索，思考，恐喝，恐怖，悲憤，悲觀，惜敗，慘敗，想像，想定，想起，意図，意識，愛好，愛着，感動，感得，感心，感服，感銘，慢心，慰勞，慶祝，慶賀，憂慮，憎惡，憤怒，憤慨，憧憬，憶測，懇請，懇願，懷古，懷柔，懷疑，懸念，懺悔，我慢，戰勝，戰慄，承服，承知，把握，投了，抗戰，拘泥，拷問，持久，指名，指定，挫折，採決，推定，推察，推測，推理，推計，推論，推量，損傷，撰生，攪乱，改心，敗北，敗戰，敗訴，敗退，敬愛，敬服，料簡，断定，断言，早とちり，早合点，是認，曲解，服從，查定，格付け，検討，構想，横着，欲求，欺瞞，歡喜，比較，気疲れ，求刑，求愛，求職，決定，決心，決意，決断，決裁，決議，油断，泣き寝入り，泥酔，深呼吸，混乱，混迷，満喫，満悦，準拠，演繹，激怒，激高，焦燥，熟慮，熟考，特約，狂信，狂喜，狼狽，猛省，玉碎，用心，申請，画定，画策，畏怖，畏敬，疲労，疲弊，痛感，発想，発覚，盲從，直感，看破，看過，眩惑，着目，督促，知覚，研鑽，確信，確定，確約，祝福，祝賀，空想，立腹，競合，策定，策謀，算段，精進，約定，約束，結論，締約，締結，羨望，習熟，考察，考慮，考案，考証，背任，脅迫，自律，自愛，自省，自覚，自認，自責，苦惱，苦悶，苦慮，苦渋，苦行，蔑視，裁可，裁定，裁決，裁量，要求，要請，規制，規定，觀念，解脱，解釈，計画，設定，許可，許容，許諾，評定，認可，認定，認容，認知，認識，誓約，説教，請求，請願，諮問，講評，謝絶，謝罪，謳歌，議決，質疑，赦免，躊躇，輕侮，輕視，辛勝，辛抱，迎撃，迫害，迷走，追從，追悼，追想，追憶，追求，追認，逆上，逆算，造反，連勝，連想，連敗，道楽，達觀，遭難，邁進，邪推，配慮，酩酊，重視，鑑別，長考，限定，陳謝，陶醉，離反，難儀，難渋，震撼，露見，頓挫，頓着，顧慮，食傷，驚嘆，驚愕

## 形容

てんでこ舞い，ガイド，マッチ，リンク，リード，一致，三唱，不自由，不足，乖離，乱立，交際，介在，伯仲，位置，併存，依拠，保養，傑出，優越，充実，充足，共通，兼任，兼備，兼務，兼営，兼業，兼用，兼職，内在，冷遇，分布，卓越，受難，台頭，合致，唱和，在学，在宅，在校，在籍，均衡，嬉し泣き，存続，実在，実存，寄与，密着，寒々，山積み，常設，常駐，平衡，幸い，広々，底冷え，引率，強調，後続，御無沙汰，念仏，息切れ，惚れ惚れ，感激，所属，才覚，払拭，拮抗，持續，撤去，撲滅，整合，断続，断線，案内，欠乏，欠落，永続，混在，添加，満載，灼熱，点在，照応，相似，相応，相違，相関，矛盾，確執，突出，立ち往生，立脚，符合，精通，紛失，自負，裏打ち，該当，調和，貢献，貧乏，起因，跋扈，近似，連係，連結，連続，連行，適合，酷似，重宝，長居，

長続き，隣接，雑居，雑踏，青々，類似

## 動作，変化

アップ，アプローチ，カンパ，カーブ，カール，キープ，タックル，ノミネート，バウンド，上昇，上陸，任命，伸縮，保温，値上がり，値上げ，免除，公募，募金，右折，向上，吸着，告別，喝采，土葬，埋葬，埋設，墮落，墜落，失墜，奏功，寄付，密封，密葬，密閉，封鎖，就任，就職，屈折，屈曲，山登り，左折，急騰，感染，打ち水，拳手，授与，授乳，接近，控除，掲揚，撤廃，散会，散水，昇格，昇給，暴騰，曲折，格上げ，標榜，歪曲，決別，決裂，没落，消化，混合，湾曲，滑落，激突，炸裂，焼香，照射，生還，登壇，登山，登用，直下，維持，緩衝，罹災，肥満，脱出，脱却，落下，落城，落石，落選，落雷，落馬，蓄財，蘇生，衝突，被覆，解散，貯水，貯留，貯金，賃上げ，転落，近接，追突，退位，退去，進級，進駐，重複，鎖国，閉塞，閉山，閉鎖，閉門，除外，除草，陥落，難航，駐屯，駐留

## 動作，感覚・感情

おんぶ，オーダー，クローズアップ，コントロール，スカウト，タイプ，ミス，リリース，一任，一括，一読，一貫，主宰，主導，乱読，交信，交流，仕返し，付託，代読，例示，依頼，係争，係留，信任，信託，傷害，充当，先遣，免税，免責，入荷，入金，公訴，再燃，再版，再編，冷却，冷蔵，出兵，出品，出展，出店，出火，出版，出稿，出資，出題，分泌，列举，列記，判示，判読，制裁，割譲，創案，加點，勉励，勉学，勉強，動員，勘定，勧誘，包括，参拝，反抗，反撃，反攻，反発，反目，反逆，反響，収容，収用，収監，収蔵，取り引き，取得，号令，合評，同封，吸収，告訴，哀願，品定め，品評，嘆息，囑託，回避，堅持，報復，外注，大騒ぎ，失策，奮励，奮発，奮起，委任，委嘱，委託，委譲，学問，学習，実演，導入，展示，巡業，巡行，帰納，廃棄，延焼，弁護，得点，復習，復讐，循環，徹夜，徹底，応戦，急報，急行，急迫，急造，急進，悪戯，悶着，愛育，愛読，慙愧，慨嘆，懲戒，扇動，打破，批判，批評，投入，投函，投擲，投棄，投獄，投球，投石，抗争，折衷，抱擁，拉致，招待，拝読，抛出，拾い読み，指揮，挑発，挿入，捻出，授賞，排便，排出，排尿，排斥，排気，排水，排泄，採点，探求，控訴，提供，提出，提案，提訴，提起，揶揄，搬入，擁立，支度，支持，放映，放棄，放流，放牧，放電，敗走，敬遠，敵視，更生，朗読，棄却，棚卸し，検温，摸索，求婚，注入，注文，注水，派出，派遣，浸水，淘汰，混用，混線，準備，激励，激論，炎上，照合，熟読，燃焼，特

注，狙い撃ち，狙撃，独学，独習，甘受，留保，発令，発声，発奮，発布，発揮，発散，  
 発案，発注，発煙，発熱，発車，発露，監督，看護，督励，研修，破棄，確保，礼賛，移  
 入，競争，競走，算出，算定，管理，精励，精読，納入，経験，給油，統一，統制，統合，  
 統括，総合，総括，総称，総計，羅列，脱帽，脱皮，自習，苦学，苦笑，苦笑い，荷揚げ，  
 薫陶，融資，装填，要約，解説，訓読，訓読み，訴訟，詐称，詠嘆，試算，誘導，誘引，  
 誘拐，誘致，読書，読経，読解，調節，諍い，謙抑，謙遜，謝恩，警戒，讓位，讓歩，負  
 託，貫通，購読，走り書き，起案，起訴，踏襲，輸血，迎合，追放，送別，逆恨み，逆襲，  
 通読，遍歴，過信，違反，違約，違背，遡及，遺棄，配列，配膳，配色，降参，除籍，陳  
 列，隱遁，集大成，震動，露出，音読，音読み，風刺，飽食，黙認，黙読，鼓舞

## 動作，形容

オン，シフト，ジョイント，プリント，マスター，リレー，享有，付与，体得，作動，併  
 設，保持，保有，保証，修得，充滿，充電，公有，共存，具備，内包，割拠，力添え，加  
 勢，助力，協力，協調，協賛，占拠，占有，同伴，同席，含有，含蓄，味付け，固定，在  
 任，在位，在勤，在職，変動，妊娠，存在，完備，実装，専有，尾行，局在，差別，常備，  
 常勤，引用，微動，愛蔵，感応，懐中，懐妊，所持，所有，所蔵，持参，掌握，推移，携  
 帯，携行，放火，整列，暖房，気乗り，活動，流通，混成，渡航，滞在，漏水，漏洩，漏  
 電，点火，焚き火，独占，現有，産出，留意，発動，相伴，私有，稼働，精勤，糊付け，  
 統率，縦走，習得，肩代わり，胎動，自営，蛇行，行列，行動，補佐，記載，躍進，身動  
 き，迂回，連動，連携，連覇，運動，運行，配備，鎮座，阻害，領有，駆動，鳴動

## 変化，感覚・感情

キャンセル，サービス，パス，乱心，仲裁，休戦，休止，倒産，傷心，全廃，冠水，冷  
 凍，凍死，合点，合祀，和解，品切れ，回生，失効，失脚，容赦，密約，帰趨，引退，得  
 心，感知，憤死，撃沈，撤回，断念，断水，根絶，棄権，横行，死傷，水没，氷結，決着，  
 沈下，沈没，沈降，潜伏，特定，狂乱，献身，理解，発狂，着眼，破産，禁煙，禁酒，結  
 審，肉薄，脱水，膠着，自制，自粛，自重，蔓延，蜂起，解約，講和，転覆，退会，退職，  
 遅配，遠慮，閉校，陥没，隠居，難破，霧散，高揚，鬱屈，鬱積，麻痺，興奮

## 変化，形容

ドッキング，一本立ち，付着，再生，剥離，半焼，即売，完売，感光，成立，承継，拡充，接合，接着，日焼け，武装，沈静，浮上，浮動，浮揚，浮遊，混雑，渋滞，焼死，癒着，直結，相続，破裂，粘着，結束，継承，背伸び，脱線，襲名，解離，軽装，透過，逸脱，適応，離党，離脱，類焼

## 感覚・感情，形容

エキサイト，コミット，一望，一苦労，傾注，切望，切磋琢磨，囑望，困窮，定着，実現，展望，希望，希求，影響，後援，志望，意気投合，懇望，成就，所望，所管，所轄，期待，渴望，熟知，熱望，直轄，祈願，窮乏，窮迫，管轄，肩入れ，自慢，苦心，要望，見栄え，象徴，連関，関与，関係，関知，関連，願望，高望み，鼎立

## 動作，変化，感覚・感情

亡命，介入，付随，侵入，侵攻，便秘，修了，偽証，入会，入信，入国，入園，入城，入場，入学，入室，入寮，入浴，入港，入湯，入社，入部，入閣，入院，入館，共演，内出血，凍結，出入り，出国，出土，出場，出塁，出席，出庫，出所，出撃，出水，出漁，出演，出血，出馬，別掲，即位，即発，参入，噴火，在中，媒介，家出，寄生，就学，就業，師事，引火，引責，従軍，復刻，打開，断食，服役，流入，流出，流失，溺死，点灯，発効，発火，発症，発芽，登場，着色，突入，競演，算入，絶食，続出，網羅，萌芽，行水，言及，読破，転入，輩出，退出，退場，退室，退廷，退社，進入，進出，進攻，運航，遮音，降臨，隨行，露呈

## 動作，変化，形容

エスカレート，スイッチ，スリップ，ハッスル，パンク，ヒット，一変，一転，上気，上達，下達，中座，中継，仲介，仲違い，伝導，伝承，伝播，伝来，伝染，伝言，伝道，伝達，体现，倍加，傾倒，元服，再現，凝縮，出世，出家，分化，分散，分断，分裂，分離，劣化，千变万化，同化，同室，命中，和睦，團結，固化，塩化，変化，変容，変形，変心，変換，変更，変異，変節，変色，変貌，変質，変遷，大成，失跡，失踪，失速，好転，孤立，尽力，帰化，座礁，心機一転，心酔，急変，急転，悪化，意味，成人，接客，改定，改訂，散逸，敵対，昇進，暗示，暗転，栄転，栄進，栄達，梅雨入り，気化，気落ち，没

入，没頭，洗練，消沈，深化，減免，減速，激動，激化，激變，災い，炭化，熟練，熟達，熱中，熱狂，発病，白熱，的中，直射，直撃，直立，直面，破綻，硬化，示唆，立ち話，立食，紅潮，絶滅，緊張，練達，老化，肥大，脱，脱会，脱毛，脱色，脱落，脱退，腐心，腐食，自立，興隆，落伍，落胆，融合，表明，表現，衰退，被弾，複合，親和，解禁，誇張，豹変，赤面，起伏，起立，軟化，転勤，転向，転換，転校，転業，轉身，軽減，辟易，退院，逆転，連合，過熱，遺伝，離別，離散，革新，風化，飽和，高騰

## 動作，感覺・感情，形容

アルバイト，乗務，仕事，作用，共働き，労働，勤勞，勤務，勤続，執務，実働，就労，店番，従業，服務，服喪，機能，発汗，粗相

## 变化，感覺・感情，形容

偏向，偏在，全焼，内通，放心，欠如，欠損，漂泊，焼損，疎通，直通，結託，通用

## 動作，变化，感覺・感情，形容

あくせく，あっさり，こざっぱり，ちょっと，どうこう，アクセス，インタビュー，セックス，ダイヤル，チャーター，ファクス，プレー，マスタ，モニタ，一定，一目，一覽，乱売，何とか，修業，入手，公刊，加給，即，執刀，失明，帰朝，得票，手術，架橋，架設，片思い，独り占め，独り立ち，独立，獲得，珍重，監理，給食，罹患，規程，課税，誹謗，警邏，軟派，退屈，録音，開眼，除雪

## 付録 B : 評判分析の実例

評判分析実験において、

- 品詞情報のみでは誤りで、意味類型情報では正しく評判文であると判断した例
- 品詞情報、意味類型情報共通して評判文を非評判文と誤って判断した例
- 品詞情報、意味類型情報共通して非評判文を評判文と誤って判断した例
- 品詞情報のみでは正しく、意味類型情報ではと誤って評判文であると判断した例

それぞれの例について 50 件ずつ掲載する。実例中の下線部は文中の最後の用言であり、文の後につく括弧内にその意味類型を示す。

### 品詞情報のみでは誤りで、意味類型情報では正しく評判文であると判断した例

白バラのただの牛乳はあんまり味の違いがわからなかつたです。( 感覚・感情 )

値段が 500 円を 超える と反則ですね。( 形容 )

便利だ、思い立ったときに好きなものが食べられる、というメリットと同時に、せつかくの四季の感覚を忘れてしまった、という悲しさを感じる。( 感覚・感情 )

その日の間ずっとちくちくと痛んだが、おかげで小さい頃によく拾って遊んでいたことや、剥いて料理したことを鮮明に思い出すことが できた。( 変化 )

レトロで綺麗なお庭が あります。( 形容 )

結構居座れると 思います。( 感覚・感情 )

でも一度行ったら気に入るお店だと 思います。( 感覚・感情 )

笑いが 止まりませんでした。( 変化 )

京都のお店はちいさいけど、その分すごくあったかくて、店員さんとの距離がちかいのが素敵だなあ と思います。( 感覚・感情 )

ほんとに京都は学生にやさしい町だなあと 思います!! (感覚・感情)

それでも京都はほんとにいいところなので、もっともっとたくさん京都を回ってみようと思います (感覚・感情)

全て京麩と京ゆばを使った料理で献立が構成されている懐石料理を味あわせてくれる、ひと味 違ったお店です。(形容)

様々な表情を見せる京麩、京ゆばに舌鼓を打ち、いつの間にか心までゆったりと くつろいでいます。(感覚・感情)

犯罪に用いられるなどによりかなりイメージを悪化させてしまったプリペイド携帯だが、一ユーザーとしては、かなり使い度 がある。(形容)

実は、私の使っている携帯電話に最近原因不明の黒い横線が入るようになってしまった。(変化)

2年前に買い換えた時とさほど変化はありませんでしたが、結局プリペイドサービスが圧倒的に値段の上では有利だという見解に 落ち着きました。(感覚・感情)

確かにプリペイドには、いくつかの弱点が ある。(形容)

ばかげた想定かもしれませんが、仮定の話として、10年後に私がプリペイド携帯ユーザーだったために契約がおじゃんになってしまう可能性だってありえるのです。(形容)

素朴な意見で自分の中にもまだこの答えに対する違和感はありますが、現時点での答えということにしておきたいと思います。(感覚・感情)

そしてその選択は正しかったように 思えます。(可能)

なんて頑丈なんだろう!と、ただただ 驚くばかりです。(感覚・感情)

そういえば、ちょっと前にドコモと au とかでも絵文字が使えるようになった。(変化)

更にいうと、業界の話ではないけど、京大でのフォーマの電波状況が改善されて、本当に1歩入れば圏外だった法経本館でばっちり三本たつようになった。(変化)

パケ代を定額にすると実質メールもウェブも使い放題になるから、と考えて、ウェブで検索したり音楽ダウンロードしたりと好き放題してしまってこんな料金になってしまうようだ。(変化)

すごいデザインのがあってこれに 決めた!! (感覚・感情)

そこまでして手に入れた携帯だから前以上に既に愛着が わいてます w (変化)

誰かと嵐山に行ったら一度除く価値はあると思う。(感覚・感情)

同じ京都市とは思えないほどの遠さを 感じてしまう。(感覚・感情)

良いスポットなので、少しでも多くの人が訪れてくれればいいと思う。(感覚・感情)

お寺とかもいっぱい あるし。(形容)

おそらく大学が自由でオープンな雰囲気醸し出しているのもこの現象の理由のひとつ



なのではないかと考えますが、これをどうとらえていいのか自分にはよく わかり ません。

( 感覚・感情 )

でも実際行ってみて、このこじんまりした感じがいいなあって 思った。( 感覚・感情 )

なんか 落ち着くし。( 感覚・感情 )

家族で何度か訪れるうちに、次第にトリコに なりました。( 変化 )

これに 魅了されました。( 形容 )

長年もの間、都であっただけあって、まち全体に厳かな雰囲気 が流れているように 感じます。( 感覚・感情 )

住んでみて改めて 思います!! ( 感覚・感情 )

そんな気持ちを反省しつつ、訪れた人に、また来たいな、やっぱり京都は期待通り、いやそれ以上に良いまちだったなと思ってほしいし、日本中・世界中に京都大好きな人が増えていけばいいなと密かに 思ってます (\*・^・\*) ( 感覚・感情 )

しかし、2週間過ぎたころから艇速が徐々に 上がってきた。( 変化 )

でもまだまだ 2000 メートルも ある!! ( 形容 )

ボートさえあればいいと 思っている。( 感覚・感情 )

生まれて初めてこんなに悔しいものはないと 思った。( 感覚・感情 )

0.17 秒の重みがずっしりと 感じられた。( 感覚・感情 )

サッカーからは本当に多くのことを学んだと 思う。( 感覚・感情 )

いつも同じくボールを投げてそれをバットで打ったり、なんの意味もないばかばかしい野球がどんどんいやに なった。( 変化 )

でも、少し馬鹿になって気楽に生き残る人生もそんなに悪くはないと 思うんだな。。( 感覚・感情 )

プロ野球の半年間戦い続けるという特性上、前半は優勝への実感が沸かず、後半はなんだか結果が見えてくる感じがしていまいち熱が 入らない。( 変化 )

ぼくは一瞬その言葉が 信じられなかった。( 感覚・感情 )

こんな話を聞いていたら彼女がすごくうらやましく なってきた。( 変化 )

でもそんなぼくの生活に対して、彼女はある種の羨望の念を 抱いていた。( 感覚・感情 )

## 品詞情報，意味類型情報共通して評判文を非評判文と誤って判断した例

おかきやせんべいの店なのだが、これがオイシイ。

初めてバンホーテンココアを飲んだときも衝撃でしたが、これは何度飲んでも衝撃です。

(動作)

99のやつよりは業務用のが、業務用のよりは平和堂のが、平和堂のよりはセブンイレブンのホームメイドテイストが、セブイレのよりはトロピカーナですかね。

ジェイムズキッチンも毎回クーポン券がもらえるのでお勧めですよ。(動作)

栗や秋刀魚、きのこなど旬の味覚がどこのスーパーでも安く売られている。(動作)

カンフォーラに友人とランチを食べに行った際、レジ際に置いてある栗を何気なく触って、見事に刺されてしまったのである。(動作)

いっぱい行ってみたいなー(動作)

最近私のはまって週に3回も行ってしまったのが、DEMOMODEっていうお店 (動作)

あとは boogaloo cafe 'もお洒落。

ホットペーパーを使ったら半額近くでディナー 食べれた (動作)

うーん、あと、だいぶ前に行った neutoron がギャラリーっぽくて素敵 やったよ (動作)

でも、一番よく行くのは中学のときも 行ってたベローチェかな (動作)

もはや '今さら 'だが、という接頭句で始めるしかないほど今さらだが、私はプリペイド携帯をずっと 使っている。(動作)

Web が使えない、通話料がやけに高いので電話をわざわざ公衆電話からかけたりする、携帯会社が謳うオトクげなサービスをほとんど 受けられない...(動作)

一方、値段の方は確かに年間最大1万8千円程度の開きがあるとは言え、現実的にお金にそうこまっているわけでもないの、ある意味 '瑣末な 'ことであつたり します。(動作)

それでも彼は今も、データも飛ばずに元気で やっています。(動作)

音楽ダウンロードといえば、au は内臓されている専用のミュージックプレーヤーで音楽を聴こうと思ったら、音楽のデータをサイトから購入するか CD から 取り込まねばならない。(動作)

結局何も変わらんの ちゃう??(動作)

なんと5年もの間 使い続けた。(動作)

しかしさすがに電池の持ちが悪くなってきたのと、たまたまキャンペーンをやっていて無料で機種変できるみたいだったから愛着のわいた携帯を手放すことに した。(動作)

しかもロゴが softbank!!

なんと品切れ。

教えてください!!! (動作)

乗車率 90%程度だろうか。

「任天堂」、「百人一首」どちらの単語も僕の心を くすぐる。(動作)

現代技術の精を尽くしたお遊びという印象を深く 受けた。(動作)

そのほかに、競技百人一首を体験できる場所もあって、2回の挑戦の後、パーフェクトを出したりして 遊んでいた。(動作)

展示としては、昔々の百人一首や大戦下の軍国百人一首が飾られていてファンとしては、眺めているだけで幸せといったところだろうか。(動作)

その名の通り、マンガンの鉱山跡を売りにしたスポットであり、その歴史案内や採掘現場である坑内の見学、飯場の模型展示などが 行われている。(動作)

戦時中においては、動員された朝鮮人労働者が多く大変な採掘作業に従事していたというこの鉱山の歴史性から、人権教育の研修の一環として訪れる団体も多いという。(動作)

その一方で、鉱物の研究者なども見学に訪れていたり する。(動作)

予想通り外国人と修学旅行生ばかりでした。

～ 京都大好き ～

そんな、素敵なまち・京都は、人をひきつける力が強いですから、多くの観光客の方が 訪れますよね。(動作)

京都の美しい光景、しっとり歩く京の人々、流れるBGMは「SoundOfMusic」の「MyFavoriteThings」。(動作)

ゆったりとした雰囲気、京の美しい景色、落ち着いた音楽、その流れで現れる文字「そうだ、京都、行こう。」(動作)

銀閣寺の近くでバイトをしていたこともあるが、ついで銀閣寺には 行かなかった。(動作)

今の1回生では1番漕暦が浅いのに持ち前の体力と精神力でレース出場権を手に した。(動作)

でも相手は日本の中でも女子クオドの強豪、東京大学。

今年のインカレで男子エイトが準優勝するくらい、男子は部員も多く、練習もきついの、私たちは彼らよりも多くの練習を こなしてきた。(動作)

朝から大好きな倅田来未をずーっと 聞いて精神統一。(動作)

先輩たちが気を 使ってくれている。(動作)

水上へ出るとき、ボート部の皆が蹴り出しに きてくれた。(動作)

ものすごく緊張しているわりには安定している。(動作)

ちゃんと詰めてる!!(動作)

もっと出さなきゃ…。(動作)

これは絶対いける!!(動作)

勝利の凱旋ではボート部の皆が喜びと驚きでいっぱいといった顔で「おめでとう!!本当に  
おめでとう!!」と言ってくれた。(動作)

本当に皆、ありがとう。

油断禁物。

## 品詞情報，意味類型情報共通して非評判文を評判文と誤って判断した例

不思議です。(形容)

ちなみに、オレンジジュースはどこのがお 好き ですか?(形容)

どんな食べ物でも大体簡単に手に入る一方、これっていつが旬だったっけ?と思うような食べ物も 少なく ない。(形容)

よい 体験... だったのであろうか?(形容)

まだ 4,5 回しか行ったこと ない けど。(形容)

\*これは すご ーく個人的な評価なので …(形容)

っていうか、高校生のときは、特に高校 3 年生のときは、食べてるとき意外ほっとするとき なかった もんなー。(形容)

かなしい かな、わたしはマンションで 1 人暮らしだったのさー。(形容)

テレビも なかった しね。(形容)

ただ、ペペロンチーノでは なかった な (笑) (形容)

だからと言って、誰かに作ってあげたいわけでも ない。(形容)

検証するには、とりあえず一緒食事をしたり、料理を作ってくれる相手が 必要 なようだ。(形容)

またラーメン屋も北区では非常に 多い。(形容)

モノの扱いが雑な自分にとっては、精密機械である携帯はできるだけ頑丈なほうが いい、という発想からの選択でした。(形容)

請求書によれば通話料金は無料通話分より少し 高い ぐらいだけど、パケット代が 10 万円分くらい。(形容)

でも定額料金そのものが 高い …(形容)

で、折角変えるならまた長く使えるのが いい じゃない?(形容)

友達と「～時に～で待ち合わせね」と約束して、一度別れてしまえば連絡手段は なし。(形容)

私は携帯電話を買うのが 遅 かった。(形容)

最近 SoftBank も 13 機種だったか増えたそうなので機種変を考えてるけどどれが いい ものかね～(形容)

当たり前のものが、ない。(形容)

実際画像は いい です。(形容)

音楽が聴きたいなら iPod 他ポータブルプレーヤーを持ち歩けばいいし、本格的な写真が  
取りたいならデジカメを持ち歩けばいい。(形容)

ゲームだって携帯機でやればいい。(形容)

そうでなくても下宿に携帯が鎮座していても意味がない。(形容)

大学は休講が多い(形容)

練習の他にも様々なサブイベントが(形容)

最初に目に入るのは、「許諾なき写真禁止」のサインである。(形容)

おそらく日本人より外国人の方が多かったのではないか。(形容)

時間がとれたころには散っていた、なんてことにならなければいいのですが。(形容)

まあええか。(形容)

あたりまえだけどきんぴかだったー!!(形容)

何といっても、他の県にはない『独特な雰囲気』!!(形容)

年に一度か二度寺社仏閣を訪ればいいところである。(形容)

とどのつまり、大事なのは心持ちということなんである。(形容)

その差はえらく小さい。(形容)

滝といっても2メートルくらいの高さの管からちょろちょろと水が流れてるようなかわいい  
もん。(形容)

...普通に重たかった。(形容)

私の頭の中にはその2文字しかなかった。(形容)

やや肌寒いくらい。(形容)

日ごろの0.17秒なんて何にもできないくらいの短さ。(形容)

それでいいってこと。(形容)

ぼくの持論では普通の人間が同時にできることは三つしかない。(形容)

個人的には悪くもないけど良くもない的中だった。(形容)

ぼくは幹部でもないし、弓道と一生付き合っていくつもりもない。(形容)

なぎなた、と聞いてスポーツとしてピンと来る人は少ないだろう。(形容)

勝つには15~20、ブルを全部三回当てるか、終わったときに点数が高ければよい。(形  
容)

この二つは周囲の環境のお陰で、熱いと感じることはないからだ。(形容)

国内ルール:嘉納先生が作った柔道本来の形に近いルール。(形容)

世間は冷たいんだなあと。(形容)

## 品詞情報のみでは正しく、意味類型情報ではと誤って評判文であると判断した例

のれんをくぐると小さな庭があり、その先に町屋風の店内がある。(形容)

割れたりして商品にならない製品のまじった「無選別」販売があるのだ。(形容)

なんでこんなに味わいが違うのでしょうか。(形容)

最近おなかが減って仕方ありませんか。(形容)

仕方ありませんね。(形容)

私はいつも昼食をとるにあたって、いかに安く大量にバランスよく食うか、しか考えて  
ません。(感覚・感情)

余談であるが、私はつい最近「秋」をビビッドに感じる体験をした。(形容)

大学のすぐ近くだけど学生さんはあんまりいないかな…(形容)

場所:百万遍を少し北に行ったら右側にある(形容)

の基準も毎回変わるかも() (変化)

三店舗行ったことがあるけど、支店こんなにあるんだーって感じ。(形容)

多分違う(形容)

あり…?(形容)

画面が見えなくなる前に(変化)

ここで私は2つの選択肢を迫られました。(感覚・感情)

月1500円~2000円の現状維持か、月3000円超でフツウの携帯ユーザーの仲間入りを  
するかです。(形容)

結局このような結論に落ち着きました。(感覚・感情)

加えて、プリペイドユーザーというものが持つ社会に対するマイナスイメージも考えて  
みました。(感覚・感情)

これは正直言ってよく分かりません。(感覚・感情)

では結局決めては何だったのか。(感覚・感情)

半年ほど前に携帯電話の機種変更をしました。(形容)

最近携帯電話業界の動きが激しいように感じる。(感覚・感情)

ボーダフォンがソフトバンクになったり、番号そのまま会社を変えることができるよう  
になったり。(変化)

略称忘れた…(感覚・感情)

これは吉田山のどこかで電波塔が新しく建ったかららしいという噂を聞いたが、詳しい理

由はよく知らない。。(感覚・感情)

話を戻すと、携帯業界、これからどうなるんだらう。(形容)

こんな日がつきに一日くらいあってもいいかと思う。(感覚・感情)

8月に嵐山へドクターフィッシュ体験で行った時に残念ながら閉館していたのでいつか行こうと思っていた場所である。(感覚・感情)

JR二条からの道は自転車・歩行者道がやたらと大きくとってあり、昔路面電車が走っていたことを思わせる。(感覚・感情)

路面電車は昔の世田谷線の雰囲気を思い出させるものだった。(感覚・感情)

都市部で通勤化されていない路面電車は珍しいといえるだろうか。(可能)

もちろん京福は観光用だろうし、実際明らかに観光客という乗客をちらほら見かけた。(変化)

嵐山に着いてからは足早にお目当ての時雨殿へ。(変化)

細い路地を通り、洒落た料亭の向こうにそれはある。(形容)

時雨殿のオフィシャルなデータについては下のリンクを辿ってもらえればと思うが、任天堂資本の百人一首博物館といったところだろうか。(感覚・感情)

中に入ると着物姿のスタッフの方がちらほら。(変化)

京都市中心部から車で1時間、旧京北町に丹波マンガン記念館という施設がある。(形容)

教養的な関心ではなく、純粋に洞窟の雰囲気が好きな人なども来ているのかももしれない。(感覚・感情)

出町柳の駅前にカリヨンというレンタサイクルがある。(形容)

2回生の中に混じっての初レース。(形容)

はじめは本当に前途多難で全く艇速が上がらなかった。(変化)

私は生意気にもそう思っていた。(感覚・感情)

そんな生意気なこと、口には出せなかったけど。(可能)

天気は曇り。(変化)

私はあまりの緊張で一言も喋れない状態。(可能)

私たちは水上でのウォーミングアップのために出艇準備。(変化)

同回生の子から「水中もっとだせー!!」という声が聞こえる。(形容)

でも確実に東大と並んでる。(形容)

いや、ちょっとでているかも?(変化)

東大が私の目の前にいる。(形容)



## 付録 C : 本論文に関する原著論文

# 動詞意味類型の曖昧性解消に向けた格フレーム情報との関連調査

岡田 正平 山本 和英

長岡技術科学大学 電気系

{okada, yamamoto}@jnlp.org

## 1 序論

日本語における用言は動詞と形容詞に分類される<sup>1</sup>。一般に動詞はウの韻で終わり、外的に観測可能な人や物の動きや変化を表すとされている。しかし動詞の中には「優れる」のような物事の特徴を表す語や「むかつく」のような人の感情を表す語が存在し、形態的な分類と意味的な分類は一致していない。

そこで中山ら [1] は形態上の特徴に関係なく、その意味によって用言を「作用性用言」と「形状性用言」に分類することを提案した。そのうえで、分類の前段階として動詞、形容詞と作用性用言、形状性用言の中間概念と言える4種類の意味類型（動作、変化、感覚・感情、形容）を定義し、人手でIPA品詞体系日本語辞書(1)（以下、IPA辞書）の動詞に付与した。この作業では動詞の持ち得る意味を落とすことの無いよう、無理に一意に決めようとせず、関係すると思われる意味類型全てを付与している。

動詞が多義性を持つ場合にも複数の意味類型が付与されることがあるが、この場合は前後の文脈から意味類型を決定できると考える。そこで本研究では多義性によって複数の意味類型が候補として付与された動詞に対して、文脈中での意味類型の推定を行うことを目的とする。本研究は、形態素解析の結果によって得られる情報のみを用いて推定を行うことを想定しており、形態素解析結果に直接情報を付加することができると考えている。形態素解析はあらゆる言語処理において基本となる処理であり、その波及効果は動詞・形容詞という品詞情報を用いる全ての処理に及ぶと考えられる。近年、自然言語処理技術の発達により、テキストマイニングツールなどの形で専門分野外の間でもその技術の恩恵に与ることができるようになってきた。そのような現状において、各種ツールの処理プロセス中に組み込むことが出来れば、専門知識を必要とせずにより意味に直結した情報を扱えるようになり、その

意義は大きい。

本稿では、予備実験として動詞の意味類型の違いによる格フレームの変化の観察を行った。

## 2 中山らの意味類型

中山らは用言の形態的な分類と意味的な分類が一致しない問題に対して、用言の新しい分類法として「作用性用言」と「形状性用言」を提案し、その分類における問題と課題を検討した。形状性用言と作用性用言はイの韻、ウの韻といった形態的な分類にはこだわらず純粋に意味によって分類される。例えば、動詞「優れる」は人や物の性質を表すため形状性用言となる。また、動詞「走る」は通常作用性用言であるが、「虫唾が走る」という1つの表現は形状性用言となる。

この分類を行う上で中山は「動詞については動作動詞や変化動詞など外的に現れる動きや、心の動きなど内的な活動で客観的に捉えることの出来ない表現が含まれている。さらには『優れる』など性質を表すものまで含まれているため、分類の前処理として各動詞が持つ意味をある程度まとめる必要がある」と述べたうえで、動作、変化、感覚・感情、形容という4種類の意味類型を定義した。この意味類型は1動詞に対して1つの意味類型が必ず決定するものではなく、複数の意味類型をもつことを許容している。以下に各意味類型の定義を示す。

**動作** 客観的に観測者が捉えることのできる運動で、その運動が終了すると運動前の状態に戻り、結果状態を表さない動詞を意味類型の”動作”とする。例えば「泳ぐ」や「食べる」などが”動作”である。

**変化** 主体に現れる運動の結果状態を表す動詞を意味類型の”変化”とする。主体が意思を持たず結果のみを表し、かつ運動が終了しても運動前の状態には戻ら

<sup>1</sup>本稿では特に言及しない限り「形容詞」に形容動詞を含めて考える

ない表現である。例えば「乾く」や「死ぬ」などが”変化”である。

感覚・感情 目, 耳, 皮膚といった感覚器官の活動と, 頭脳や心の働きなどを表す知情意を意味類型の”感覚・感情”とする。定義では, 意思性のある感覚器官の活動を”動作”とし, 「見える」や「感じる」など非意志的なものを感覚とするとしている。また「愛する」など, 心の動きだけでなく客観的に観察可能な動きの意味も含んでいる動詞に対して, “感覚・感情”の意味類型だけでなく”動作”も持つと定義している。

形容 人や物の様子や性質, 形, 存在, 関係を表す表現を意味類型の”形容”とする。“形容”は一般的な形態上の分類である形容詞の集合と, 意味的な分類の上で同義になると考えられている。

”動作”または”変化”の動詞は作用性用言に, “形容”である動詞は形状性用言に漏れ無く対応する。

中山らは IPA 辞書中の各動詞 (12,648 表現) に対して人手で意味類型を付与した。この際, 動詞が持ち得る意味を落とすことの無いよう, 関係すると思われる意味類型全てを付与した。

### 3 意味類型の自動推定に向けた調査

中山らが意味類型を付与した動詞の中には, 動詞の多義性によって複数の意味類型が付与されたものがある。これらは, 前後の文脈が決定すれば意味類型も決定できると考える。例えば, 動詞「満たす」は”動作”と”形容”の意味類型が付与されているが「コップに水を満たす」という文中ではコップに水を注ぎ入れ満杯にするという運動を表すので意味類型は”動作”, 「条件を満たす」という文中では条件に適合するという状態や性質を表すので”形容”であると言える。そこで本研究では, 多義性によって複数の意味類型が候補として付与された動詞に対して, 文脈中での意味類型の推定を行うことを目的とする<sup>2</sup>。

本稿では意味類型推定のための予備調査として, 意味類型の違いによって動詞の文中での使われ方が異なるのかを調べる。そこで, 単一の意味類型を付与された動詞を対象に, 京都大学格フレーム (2) (以下, 格フレーム辞書) を用いて各々の意味類型が付与された動詞が取る表層格および項について調べた。格フレー

<sup>2</sup>ただし複数の意味類型が付与された理由のアンノテーションはされていない。そのため, あくまで多義性により複数意味類型が付与されたと思われる動詞について推定を行うものとする。

表 1: 単一の意味類型が付与された動詞数 (分母) および格フレーム辞書中に共通で出現する動詞数 (分子)

動作	変化	感覚・感情	形容
1516/6637	418/1530	360/1441	78/358

ム辞書は JUMAN 辞書に基づいており, IPA 辞書とでは単語の単位が異なるため IPA 辞書の動詞の表記と格フレーム辞書の述語 (動詞) の表記が一致したもののみを対象とした。この際, 態の違いは考慮していない。また, 格フレーム辞書中では複数形態素での表現も登録されているが, その場合は IPA 辞書の動詞が最初の形態素と一致したものを対象とした。単一の意味類型が付与された動詞数および, その中で格フレーム辞書中にも出現した動詞数は表 1 の通りである<sup>3</sup>。

#### 3.1 表層格の頻度

意味類型ごとに取りやすい表層格に違いがあると仮定し, それぞれ格フレーム辞書中の 9 種の表層格の出現頻度を累計した。また, 複数の意味類型が付与されている動詞が文中においてどの意味類型になるかは一様な確率であると仮定し, 表層格ごとにどの意味類型の動詞に取られやすいかを集計した。ただし単一の意味類型が付与された動詞全てを対象に集計すると, “動作” が付与された動詞の絶対数が多いため偏りが生じてしまう。そこでランダムサンプリングにより意味類型ごとの動詞数を, 最も動詞数の少なかった”形容”と同じ 78 に揃えた上で表層格ごとの意味類型の出現頻度を累計した。

各意味類型の表層格の出現頻度を表 2 に, その内訳を図 1 に示す。サンプリングした動詞の表層格の出現頻度を表 3 に, 全動詞の場合とを比べた各意味類型の表層格出現頻度の相関係数を表 4 に示す。いずれの意味類型でも 0.9 以上の相関係数が得られたので, このサンプリングは妥当であると言える。サンプリング後の表層格ごとの動詞の意味類型の割合を図 2 に示す。

ガ格とト格を取る動詞は, ”形容” が付与されたものが 7 割超存在し大きな割合を占めている。”形容” が付与された動詞は, 主語となる名詞とセットでその名詞状態を説明する働きがあるためガ格を取りやすく, また文中で省略されにくい。一方でその他の意味類型が付与された動詞については, 主語が既に話題に挙がっ

<sup>3</sup>同じ語の漢字表記とひらがな表記に関して IPA 辞書では別のエントリであるのに対して, 格フレーム辞書中では同一のエントリにまとめられていることに注意されたい。共通の動詞の数が少ないように見えるのはこのことが大きな原因の一つである。

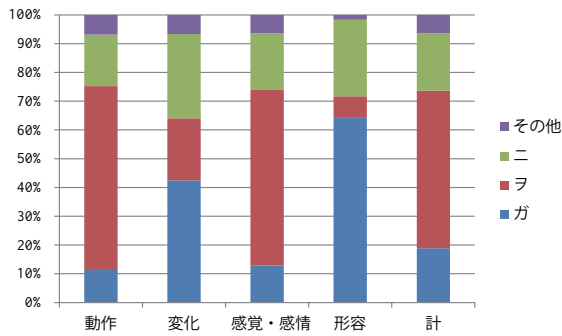


図 1: 動詞の意味類型ごとの動詞の取る表層格の割合

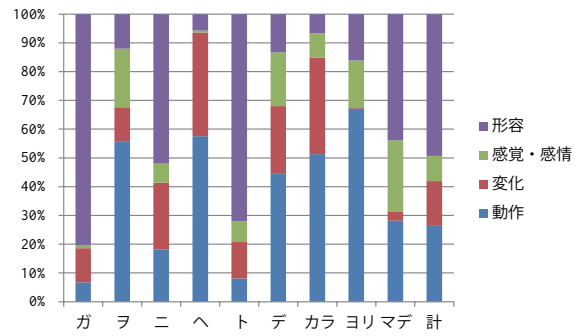


図 2: サンプルング後の動詞に対する表層格ごとの取られる動詞の意味タイプの割合

表 2: 全動詞における意味類型ごとの総表層格数

動作	変化	感覚・感情	形容
2億 2664 万	3022 万	2278 万	2424 万

表 3: サンプルング後の意味類型ごとの表層格数

動作	変化	感覚・感情	形容
1296 万	764 万	432 万	2424 万

ている場合に主語が省略される場合が多く、その他の情報を説明することが多いため、ガ格以外の表層格を取るのだと考える（例えば「私が食べる」というより「りんごを食べる」という用法の方が多い）。ト格については、「形容」以外の意味類型が付与された動詞については「ゆっくりと」などと言った副詞的な用法が多いが、「形容」が付与された動詞はそれに加えて「異なる」や「似る」のような動詞の比較対象を表すためによりト格を取りやすいのだと考える。

ヲ格、ヘ格、カラ格、ヨリ格については、「動作」が付与された動詞が 5 割超存在し、大きな割合を占めている。「動作」が付与された動詞は、動詞全体に対する割合が大きく多様な動詞を含んでいるため多くの表層格を取るのだと考える。これらの表層格の情報は、「形容」と「動作」の 2 つが付与された動詞の意味類型推定を行う場合に、ひとつの大きな手がかりになる。一方で、「変化」や「感覚・感情」は表層格ごとに割合の「変化」はあるが、全体で大きな割合を占める表層格は無かった。したがって、{ 形容 | 動作 } と { 変化 | 感覚・感情 } の間で判別が必要な場合には表層格の種類の情報には推定に貢献しにくいと考える。

表 4: 全動詞とサンプルングされた動詞間での、意味類型ごとの表層格頻度の相関係数

動作	変化	感覚・感情	形容
.998	.942	.989	1.000

### 3.2 項の種類

表層格ごとにその意味類型を持つ動詞にしか取られない項の名詞意味属性がその意味類型を特徴付けると仮定し、IPA 辞書と格フレーム辞書に共通に出現した全動詞を対象に意味類型固有の表層格と名詞意味属性のペアを収集した。名詞意味属性には、日本語語彙大系 (3) の名詞・固有名詞意味属性（以下、まとめて名詞属性）を利用した。

IPA 辞書と格フレーム辞書に共通な動詞全てを対象にして、意味類型ごとに得られたその意味類型を持つ動詞にしか取られない項の名詞属性の数を表 5 に示す。また、得られた名詞属性中から著者 1 人の主観により、比較的抽象度の高いと思われるものの一部を以下に示す。

動作 カラ格:事象, 性質, 具体, ヘ格:精神, 具体, ト格:事象, 性質, 主体, ガ格:抽象, 国名, マデ格:事, 行為, 状態, 形状, 数量, 組織, 人間, デ格:表情, 性質, 衣料, ヨリ格:出来事, 地方名, 都市, ヲ格:事象

変化 ヘ格:感覚, 出来事, ト格:生命現象, マデ格:形, デ格:無生物, ヨリ格:食料

感覚・感情 ヨリ格:事, 行為, 道具

形容 ヘ格:宗教名, ト格:乗り物, マデ格:特徴, ヨリ格:形, 生物

表 5: 意味類型固有の表層格-名詞属性ペアの数

動作	変化	感覚・感情	形容
3243	196	198	137

これらの表層格-名詞属性のペアが文中に現れた場合は、意味類型推定のための決め手になると考える。実際には意味類型が4種類付与されていなければ、付与されている意味類型内で固有であればいいので、扱うことのできる情報はより多い。また、日本語語彙大系の名詞属性は木構造になっているため、得られた属性の親属性が意味類型に固有かどうかを調べていくことで、より多くの情報を得ることができる。

## 4 関連研究

本研究は、語義を4種類に規定した上での動詞の語義曖昧性解消タスクであるとみなすことができる。文中の動詞に語義を付与するシステムには、竹内ら [2] による意味役割付与システムが挙げられる。意味役割付与システムは入力文を述語項構造解析を行い、その後述語の語義の同定と項の意味役割付与を行う。言語資源を人手で構築することで、深い処理に向けた高級な情報を付与している。また本研究とは述語に付与する語義も異なっている。本研究では動詞自身の語義のみを扱っているのに対し、竹内らは動作対象の状態変化などを含むより広い範囲を扱っている。

また、動詞をその意味により分類することは、日本語教育の上でも意義がある。谷口 [3] によれば、日本語を正しく使いこなす上で動詞の種類についての理解が重要であるとし、その動詞の分類法の一つにアスペクトによる分類を挙げた。これは金田一 [4] が提唱した分類法であり、具体的には第二種の動詞（継続動詞）は中山の意味類型の”動作”と、第三種の動詞（瞬間動詞）は”変化”と、第一種の動詞（状態動詞）及び第四種（常に「テイル」を伴う動詞）は”形容”と相関が強いと考える。このことから、意味類型の自動推定を用いることで日本語の文に日本語学習者に向けた情報を付与できると考える。

## 5 まとめ

中山らは日本語の用言を意味的に分類する前段階として、4種類の意味類型を提案し、IPA 辞書の動詞について人手で付与を行った。本研究では、多義性によ

て複数の意味類型が付与された動詞について、文脈から意味類型の推定を行うことを目的とする。今回はその予備実験として、単一の意味類型が付与された動詞について、意味類型ごとにその動詞がとる表層格と、項の違いを調べた。表層格に関しては、ガ・ト格を取る動詞は意味類型の”形容”が付与された動詞が、ヲ・ヘ・カラ・ヨリ格を取る動詞は”動作”が付与された動詞が大きな割合を占めることがわかった。また、“変化”と”感覚・感情”については表層格以外の情報が必要であることがわかった。項については日本語語彙大系上の名詞属性を利用して、意味類型に固有の表層格-項の名詞属性の組を得ることができた。

今後の展望として、今回得られた結果から実際に自動推定器を作成し評価とフィードバックを繰り返しながら精度の向上を目指す。完成した推定器は公開する予定である。

## 使用した言語資源およびツール

- (1) IPA 品詞体系日本語辞書 . Ver.2.7.0 . <http://mecab.sourceforge.net/src/>
- (2) 河原大輔, 黒橋禎夫 . 京都大学格フレーム Ver 1.0 . 言語資源協会, 2009 . <http://www.gsk.or.jp/catalog/gsk2008-b>
- (3) 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 大山芳史, 林良彦 . 日本語語彙大系 CD-ROM 版 . 岩波書店, 1999 .

## 参考文献

- [1] 中山匠, 山本和英. 用言の新しい意味類型 — 作用性用言と形状性用言 — . 言語処理学会 第 17 回年次大会 発表論文集, pp. 560–563, 2011.
- [2] 竹内孔一, 土山傑, 守屋将人, 森安祐樹. 類似した動作や状況を検索するための意味役割及び動詞語義付与システムの構築. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, Vol. 109, No. 390, pp. 1–6, 2010.
- [3] 谷口秀治. 日本語教育における動詞の分類について. 大分大学留学生センター紀要 2 号, pp. 53–63, 2005.
- [4] 金田一春彦. 日本語動詞の一分類. 日本語動詞のアスペクト, pp. 5–26. むぎ書房, 1976 (原論文 1950) .

# Semantic Type Disambiguation for Japanese Verbs

Shohei Okada and Kazuhide Yamamoto  
Department of Electrical Engineering  
Nagaoka University of Technology  
Nagaoka, Japan  
{okada, yamamoto}@jnlp.org

**Abstract**—The interest has been increasing in recent years in extracting and analyzing evaluations and opinions of service or products from large bodies of text. It is important to classify predicates according to sense because whether or not a statement includes the speaker’s opinion depends strongly on its predicate. It is generally assumed that Japanese part-of-speech (POS) for predicates is classified according to sense; however, the POS classifications differ from their semantic classification. On this subject, semantic types, which aim to classify predicates, have been proposed. In this paper, we describe semantic types and present our construction of a disambiguator for Japanese verbs. Specifically, we constructed this disambiguator using a support vector machine by building feature vectors. We used semantic categories of noun and results of morphological analysis for the feature vectors. We then achieved 69.9% accuracy of disambiguation for newspaper articles using 10-fold cross-validation.

**Keywords**—Word Sense Disambiguation (WSD)

## I. INTRODUCTION

The interest in extracting and analyzing evaluations and opinions of services or products from large bodies of text has been increasing in recent years. Whether or not a statement includes the speaker’s opinion depends strongly on its predicate. Yang and Cardie [1] said “the majority of opinion expressions involve verb phrases”. Therefore, it is important to classify predicates according to sense, and Japanese predicates consist of adjectives and verbs. It is generally assumed that adjectives signify properties or states of people or objects, and while verbs signify observable actions or changes. However, there are verbs that signify properties, such as “優れる (to excel)” and emotions such as “むかつく (to get frustrated)”. Therefore, *part-of-Speech* (POS) differs from semantic classification for Japanese predicates.

In their 2011 study, Nakayama and Yamamoto [2] proposed that predicates should be classified according to sense and, accordingly, they defined four *semantic types*: *action*, *change*, *emotion*, and *modification*. According to Nakayama and Yamamoto, all adjectives are of the type *modification*, while verbs are of multiple types. They then annotated all verbs in a dictionary for type, and, in order to retain information, they annotated all possible types that could relate to a given verb. Verbs that have ambiguity of sense can be annotated multiple types. In this case, one of the annotated types can be determined from the context. Okada and Yamamoto [3] investigated the relationship between semantic types and case frame information. They

found that the combination of surface case information with the semantic categorization of a noun in a statement is an effective means to determine semantic type.

In this study, therefore, we constructed a semantic type disambiguator, that determines verb type from context based on Nakayama and Yamamoto’s classification. We supposed that the disambiguator only utilizes information from the results of morphological analysis and existing language resources. Thus, we were able to directly add type information to the results of morphological analysis like POS tags. Morphological analysis and POS tagging are fundamental operations for *Natural Language Processing* (NLP). By enabling us to consider information that more accurately reflects verb sense using semantic type instead of POS, this study can contribute to Japanese NLP which uses POS as semantic categories. For an example of opinion mining, when applying the method proposed by Scholz and Conrad [4] to Japanese, improvement of precision can be expected using semantic type categorization instead of POS.

In this paper, we first introduce Nakayama and Yamamoto’s definition of semantic types and the concept of a *Semantic Type Dictionary*. In the third section, we describe the construction of the semantic type disambiguator. We will evaluate the disambiguator in section IV, before concluding in the final section.

## II. SEMANTIC TYPES

### A. Definition of Semantic Type

In their study, Nakayama and Yamamoto proposed classifying predicates according to sense and discussed the associated problems. For example, a verb “優れる (to excel)” behaves like an adjective insofar as it signifies the property of a person or an object. A verb “走る (to run)”, which usually behaves as a verb, forms a phrase “虫唾が走る”, which means “to give someone the creeps” in English and behaves like an adjective. Nakayama and Yamamoto found that aggregation of verb senses was necessary to be preprocessed for the classification due to the diversity in sense of verbs. They then proposed that verbs have four semantic types: *action*, *change*, *emotion* and *modification*. Verbs do not conform uniquely to one type, but can have multiple types. The four types are defined as follows:

*action*: Expressions that signify objectively observable motion and in which the state does not change before and after the motion.

Table I  
ANNOTATED SEMANTIC TYPES AND NUMBER OF VERBS.

action	change	emotion	modification	# of verbs
Y	N	N	N	6637
N	Y	N	N	1531
N	N	Y	N	1441
N	N	N	Y	358
Y	Y	N	N	469
Y	N	Y	N	1127
Y	N	N	Y	190
N	Y	Y	N	396
N	Y	N	Y	113
N	N	Y	Y	72
Y	Y	Y	N	143
Y	Y	N	Y	61
Y	N	Y	Y	35
N	Y	Y	Y	42
Y	Y	Y	Y	33
total				12648

Y: annotated, N: not-annotated

e.g.) 泳ぐ (to swim), 食べる (to eat)

*change*: Expressions that signify a state as a result of a motion and in which the state after the motion differs from the state before the motion.

e.g.) 乾く (to dry), 死ぬ (to die)

*emotion*: Expressions that signify the operation of a sense organ, such as eyes, ears, and skin, or mental actions. According to Nakayama and Yamamoto, verbs that express intentional operations, such as “見る (to look)” and “考える (to think)”, have the type *action* and those that express unintentional operations, such as “見える (to see)” and “感じる (to feel)”, have the type *emotion*.

*modification*: Expressions that signify properties, shapes, beings, or relations. According to Nakayama and Yamamoto, all of adjectives are of this type.

e.g.) 優れる (to excel), 異なる (to differ)

All *action* or *change* verbs behave like verbs and *modification* verbs behave like adjectives.

### B. Semantic Type Dictionary

Nakayama and Yamamoto manually annotated the semantic types for all 12,648 verbs in the *IPADIC for Japanese*<sup>(1)</sup>. They annotated all types that can be related to a verb without losing the verb’s information. The breakdown of annotated types and number of verbs are shown in Table I. 2,681 verbs (*i.e.*, approximately 20% of all verbs) are annotated as multiple types.

## III. CONSTRUCTION OF THE SEMANTIC TYPE DISAMBIGUATOR

Some verbs are annotated as multiple types in the dictionary due to ambiguity of verb sense. We assumed that such verb types were determined from context information. For example, the verb “満たす” has both the sense of “to fill” and “to satisfy” in English. If the verb appears in the sentence like “コップに水を満たす (to fill a glass with water)” the type is *action* and “条件を満たす (to satisfy a condition)” *modification*. In this study, we aim

Table II  
FEATURES FOR SVM.

feature	# of features	value
target verb	2681	0 or 1
inflected form	19	0 or 1
semantic types	4	0 or 1
particles	118	0 or 1
semantic categories	21	real value

to disambiguate the types of verb that have ambiguity and multiple type using context information.

In their study, Okada and Yamamoto [3] investigated the relationship between semantic types and case frame information. They found that using surface frame information is insufficient on its own, and combining this information with semantic categorization of nouns in the statement is an effective way to determine semantic type. Therefore, we achieve disambiguation as a multi-class classification problem using a *support vector machine* (SVM) by building feature vectors. In the following it explains what feature has been used for the feature vector.

### A. Feature Vector

First, an input sentence, which includes a target verb, was analyzed by the morphological analyzer *MeCab*<sup>(2)</sup>. Morphological features were then generated from the results of the analysis. These features are shown in Table II. Description for each feature is as follows:

1) *target verb*: We assumed that the verb type was dependent on the verb itself. Hence, we used the target verb as a feature. The target verb takes value 1 and the other verbs take 0 for all 2,681 verbs which are annotated as multiple types in the dictionary.

2) *inflected form of target verb*: The inflected form of the verb relates to its role in the sentence. Hence the inflected form relates to verb type. Therefore, we used the inflected form of the target verb as a feature. The inflected form takes value 1 and the other forms take 0 for all 19 inflected forms in IPADIC.

3) *semantic types*: We used the types annotated to the verb in the dictionary as features. The types annotated take value 1 and the other types take 0.

4) *particles*: The case is indicated by particular particles in Japanese. We used the particles that decide the case, specifically case particle and binding particle, as features. The particles that appear in the sentence take value 1 and other particles take 0 for 118 of the particles in IPADIC.

5) *semantic categories of noun*: We used occurrence frequency of semantic categories of noun that appear in the sentence as features. For the categories of noun, we used semantic categories of noun and proper noun in *Goitaikei — A Japanese Lexicon*<sup>(3)</sup>. *Goitaikei* is a Japanese thesaurus that consists of 300,000 words and 3,000 semantic categories. The categories are used with generalization in order to avoid sparseness problem. Specifically, we limited the depth of the thesaurus of the categories from the root category. We set the depth limitation to 3 for the results of the preliminary experiment. The number of categories were reduced to 21 by the generalization. There

Table III  
FREQUENCIES OF SEMANTIC TYPE IN THE DATA SET.

action	change	emotion	modification	total
451	614	255	680	2000

are certain nouns that have multiple categories. Therefore, we formulated occurrence frequency  $freq(c)$  for semantic categories  $c$  as follows:

$$freq(c) = \sum_{n \in N_c} \frac{freq(n)}{|C_n|}, \quad (1)$$

where  $N_c$  is set of nouns that have the category  $c$ ,  $freq(n)$  is occurrence frequency of noun  $n$  in the sentence, and  $C_n$  is set of categories that the noun  $n$  has.

Conventionally, syntactic structure information is utilized in a task of *Word Sense Disambiguation* (WSD) or *Semantic Role Labeling* (e.g. Johansson and Nugues [5]), which the semantic type disambiguation is similar to. However, we use only morphological features because dependency parsers may not be robust for casual text such as blog text.

We apply the Libsvm<sup>(4)</sup> for classification. We used a linear kernel for the kernel function due to sparseness of features and adjusted the parameters to maximize accuracy.

#### IV. EVALUATION

We evaluated our disambiguator using newspaper articles and compared with two baselines. In the following it explains about the setup of evaluation and the results.

##### A. Setup

We used newspaper articles<sup>(5)</sup> published in 2004 as the data set. First, sentences that included a verb that have multiple types were extracted from the data set. These verbs were then used as the classification target. Next, 2,000 of these sentences were randomly sampled as the data set. One of the authors manually annotated the most appropriate type to each verb. The frequency of each type, as a result of annotation, is shown in Table III.

We evaluated the accuracy of the disambiguator for the data set using 10-fold cross-validation. In order to achieve this, we set following two baselines:

1) *random*: One of the types that the verb has was selected at random.

2) *frequency*: The most frequently type in the types that the verb has was selected. The priority follows the order of *modification* > *change* > *action* > *emotion* for Table III.

In this study, the disambiguator is not compared with other systems. It is because there is no other studies about disambiguation for the semantic types.

##### B. Results and Discussion

The evaluation result is shown in Table IV. We achieved 69.9% accuracy and marked better than baseline results. The breakdown of the disambiguation results is shown in Table V. According to the table, both of recall and

Table IV  
EVALUATIVE RESULT.

method	accuracy [%]
SVM	69.9
baseline1 (random)	24.1
baseline2 (frequency)	30.1

Table V  
BREAKDOWN OF THE DISAMBIGUATION RESULTS.

actual	predicted				total	recall [%]
	act	cha	emo	mod		
action	276	64	61	50	451	61.2
change	31	435	10	138	614	70.8
emotion	63	28	147	17	255	57.6
modification	23	99	18	540	680	79.4
total	393	626	236	745	2000	-
precision [%]	70.2	69.5	62.3	72.5	-	69.9

precision for *modification* mark the worst in the four types. We considered this is because the number of training data for *modification* is the least; therefore, increasing data set for *modification* or weighting based on number of the data set for each type during training can improve the disambiguation accuracy.

The achieved accuracy seems low, however, we think it will not be serious problem in applications. As we can see for Table I, approximately 80% of all verbs have single semantic type. We roughly estimate more than 90% of verbs in sentences can be labeled with a correct semantic type by the disambiguator.

Furthermore, we examined the change of accuracy by size of data set in order to evaluate whether the size was sufficient. The results are shown in Figure 1. According to the figure, because the accuracy is on an upward trend in the range, we expected higher accuracy with a larger data set.

#### V. RELATED WORK

We can assume that our study is a WSD task in which the word sense is limited to four types. WSD for verbs is a fundamental topic of NLP. In English, the classification based on two verb categorization schemes, *VerbNet* [6] and *FrameNet* [7], has been addressed. Croce et al. [8] proposed a verb classification model according to VerbNet and FrameNet. They used the similarity of two verbs based on syntactic dependencies and kernel function.

In Japanese, Takeuchi et al. [9] constructed a system that estimates verb sense in a sentence. This system conducts argument structure analysis using manually constructed language resource. In this analysis, the sense of a predicate is estimated and its semantic role is added to the argument, in contrast with our study, which only utilized shallow information.

To classify verbs according to sense contributes to Japanese language education. According to Taniguchi [10], understanding for types of verb is important to correct use of verbs. Classification by aspect proposed by Kindaichi [11] is one of the sense types. The second verb (continuing verb) in the classification relates to the type *action*, the third verb (instant verb) relates to the type



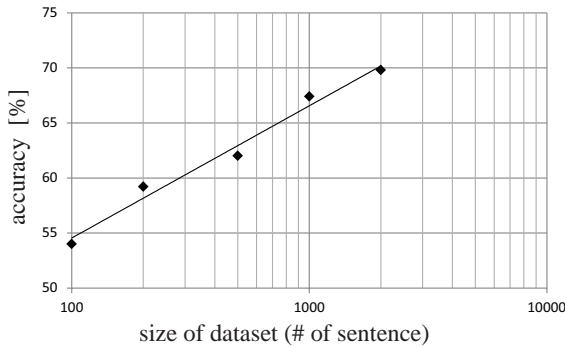


Figure 1. Relationship between size of data set and accuracy.

*change* and the first (state verb) and fourth (verb always entails “テイル”) relate to the type *modification*. Hence, the semantic type adds information for Japanese language learners.

## VI. CONCLUSION

We constructed a disambiguator for semantic types, as proposed by Nakayama and Yamamoto. This disambiguator was achieved using an SVM that utilizes information from existing resources and the results of the morphological analyzer. We then achieved a 69.9% accuracy rate for newspaper articles using 10-fold cross-validation.

In future work, we will evaluate the effectiveness of these semantic types using the semantic type dictionary and the disambiguator by applying them to a particular task such as opinion mining.

The disambiguator will be available to the public.

## TOOL AND RESOURCES

- (1) Information-technology Promotion Agency Dictionary (IPADIC) for Japanese. Version 2.7.0. <http://mecab.sourceforge.net/src/>
- (2) Morphological analyzer MeCab. Version 0.993. <http://mecab.sourceforge.net/>
- (3) Goitaikei — A Japanese Lexicon. <http://www.kecl.ntt.co.jp/icl/lirg/resources/GoiTakikei/index-en.html>
- (4) Libsvm. Version 3.18. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- (5) Nihon Keizai Shinbun (literally, Japanese Economics Newspaper) all article database CD-ROM, 2004.

## REFERENCES

- [1] B. Yang and C. Cardie, “Extracting Opinion Expressions with semi-Markov Conditional Random Fields”, In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing (EMNLP) and Computational Natural Language Learning (CoNLL)*, pp. 1335-1345, 2012.
- [2] T. Nakayama and K. Yamamoto, “New Semantic Types for Predicates”, In *Proceedings of the 17th Annual Meeting of the Association for Natural Language Processing*, pp. 560-563, 2011. (in Japanese).

- [3] S. Okada and K. Yamamoto, “Investigation of Relation with Case Frame Information for Semantic Type Disambiguation” (original Japanese), In *Proceedings of the 20th Annual Meeting of the Association for Natural Language Processing*, pp. 524-527, 2014. (in Japanese).
- [4] T. Scholz and S. Conrad, “Opinion Mining in Newspaper Articles by Entropy-based Word Connections”, In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1828-1839, 2013.
- [5] R. Johansson and P. Nugues, “Dependency-based Semantic Role Labeling of PropBank”, In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 69-78, 2008.
- [6] K. K. Schuler, “VerbNet: A broad-coverage, comprehensive verb lexicon”, *Dissertations available from ProQuest*. Paper AAI3179808. University of Pennsylvania, 2005.
- [7] C. F. Baker, C. J. Fillmore, and J. B. Lowe, “The Berkeley FrameNet Project”, In *Proceedings of the 1998 Joint Conference on 36th Annual Meeting of the Association for Computational Linguistics (ACL) and 17th International Conference on Computational Linguistics (COLING)*, pp. 86-90, 1998.
- [8] D. Croce, R. Basili, A. Moschitti, and M. Palmer, “Verb Classification using Distributional Similarity in Syntactic and Semantic Structures”, In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 263-272, 2012.
- [9] K. Takeuchi, S. Tsuchiyama, M. Moriya, and Y. Moriyasu, “Construction of Argument Structure Analyzer Toward Searching Same Situations and Actions”, *the Institute of Electronics, Information and Communication Engineers (IEICE) Technical Report. Natural Language Understanding and Models of Communication (NLC) 109(390)*. pp. 1-6. 2010. (in Japanese).
- [10] S. Taniguchi, “Verb Classification for Japanese Language Education”, *Oita University International Student Center bulletin No. 2*, pp. 53-63, 2005. (in Japanese).
- [11] H. Kindaichi, “Classification for Japanese Verbs”, *Aspect of Japanese Verb*, pp. 5-26, 1976. (in Japanese).

# 評判分析における品詞情報と意味類型情報の有効性比較

岡田 正平 山本 和英

長岡技術科学大学 電気系

{okada, yamamoto}@jnlp.org

## 1 序論

日本語における用言は動詞と形容詞に分類される<sup>1</sup>。その分類基準はその終止形が母音ウを持つ拍で終わる(動詞)、拍「い」で終わる(形容詞)という形態的なものである。また、一般的に動詞は客観的に観測可能な人や物の動きや変化を、形容詞は物事の特徴や性質を表すとされている。しかし形態的な基準に基づく品詞分類は、その品詞に属する語の意味を完全に分離しているわけではない。例えば動詞の中には「優れる」のような物事の特徴を表す語や「むかつく」のような人の感情を表す語が存在している。したがって、テキストから動作のみを抽出する目的ですべての動詞を抽出すると動作とは無関係な上記のような動詞が含まれてしまう、などの問題が生じることになる。

このような問題に対し、本研究室の先行研究において中山ら [1] は形態上の特徴に関係なく、その意味によって用言を「作用性用言」と「形状性用言」に分類することを提案した。そのうえで、分類の前段階として動詞、形容詞と作用性用言、形状性用言の中間概念と言える4種類の意味類型(動作, 変化, 感覚・感情, 形容)を定義した。また中山らは、用言のうち形容詞は常に意味類型の「形容」のみを取り、動詞はその語義や用法により同じ動詞でも異なる意味類型を取り得るとし、IPA 品詞体系日本語辞書<sup>(1)</sup>中の全ての動詞に対して、取り得る意味類型を人手により付与したSNOW D1: 動詞意味類型辞書を作成し、公開した<sup>2</sup>。

また我々[2]は、文中の動詞の意味類型はその文脈から1つに決定できると仮定し、形態素解析結果と既存の言語資源から得られる情報を素性とした教師あり機械学習を用いて文中の動詞の意味類型を1つに決定する意味類型曖昧性解消器を作成した。

これらの先行研究により形態素解析済みの日本語文中のすべての用言に意味類型を付与することができるようになった。用言に付与された意味類型情報は動詞・

形容詞という表層的な情報を用いた自然言語処理を行うタスクにおいて、品詞情報の代わりに利用することができる。応用例としては評判分析や感情解析、テキスト中からの動作や変化の抽出などが挙げられる。本研究では評判分析の例をとり意味類型の有用性を示す。具体的には、品詞の代わりに意味類型を用いることでブログ記事中から評判表現を含む文の抽出の精度を4.7ポイント、F値を27.2ポイント向上させた。

## 2 関連研究

品詞情報の代わりに意味類型情報を利用することは、語の意味をより正確に扱うことにつながり、性能向上に寄与できると考える。本研究では意味類型を評判分析に応用することでその有用性を示す。

近年、個人がWeb上に意見を発信することが可能になったことに伴い、テキストを対象とした評判分析・感情解析への関心が高まっている。品詞情報を用いている研究としては以下のような例が挙げられる。Scholzら [3] は語ごとに計算されるエントロピーに基づく指標を品詞毎に平均し素性とする手法で、新聞記事中の文に対する肯定/否定/中性(客観)の極性推定で最高65%の精度を達成している。また、小林ら [4] は意見を(対象, 属性, 評価値)という3つ組で定義、共起パターンに基づく属性・評価値表現の半自動抽出方法を提案しており、抽出時のノイズ軽減を目的としたフィルタリングの1つに品詞情報を用いている。感情解析においては、Nakagawaら [5] は依存構造木のノードの極性を表す隠れ変数と、事前極性や極性反転の有無、品詞等の組み合わせを素性として用いた条件付き確率場を用いて文の感情極性を2値分類する手法を提案し、日本語を対象に最高84.7%の精度を達成している。

品詞情報を用いない手法の例としては、Socherら [6] は再帰的ニューラルネットワークをテンソルの導入によって改良したRecursive Neural Tensor Networkによって、文レベルでの感情極性2値分類で85.4%の精度を達成している。

<sup>1</sup>本稿では特に言及しない限り形容動詞を形容詞に含めて考える。

<sup>2</sup><http://www.jnlp.org/SNOW/D1>

### 3 意味類型の評判分析への応用

中山らは日本語の用言を意味的に分類することを目的として、動作、変化、感覚・感情、形容という4種類の意味類型を定義した。文中の用言はその用言が持つ意味によって1つの意味類型を持つと考える。また、同じ語であっても文脈によって持つ意味類型が異なる場合がある。例えば「コップに水を満たす」の「満たす」は意味類型の動作を、「条件を満たす」の「満たす」は意味類型の形容を持つ。現在までの我々の先行研究により、形態素解析結果から文中の用言に対してその用言が持つ1つの意味類型を付与することが可能になっている。そこで本研究では意味類型を実際の自然言語処理タスクに応用し、その有効性を示す。

応用例として本研究では評判文抽出を扱う。評判表現の多くは「良い」「～が多い」などの形容表現が用いられる。意味類型の情報をを用いることで「～に優れる」といった動詞の形容表現を扱うことができる、また「～を感じる」「～だと思う」のような意味類型の感覚・感情に対応する表現も評判になりやすい。このようなことから、意味類型情報を利用することでより網羅性高く評判表現を扱うことができ、性能向上に貢献できるという仮説を立てた。

対象とするコーパスには249記事、4186文からなる人手による解析済みブログコーパスであるKNBコーパス<sup>(2)</sup>を用いた。本研究における評判表現の定義はKNBコーパスに従い「ある対象に対して述べられた肯定的、もしくは否定的判断や態度、叙述」とする。KNBコーパスは評判表現についてアノテーションされており、「感情」「批評」など7種の評判タイプや評価の極性、評価保持者、評価対象といった詳細な情報も付与されている。しかし、本研究の目的は純粋な品詞情報と意味類型情報の比較であるため、評判表現の詳細な情報についての予測・同定は行わない。本研究ではKNBコーパス中で評判表現を含む文を評判文、含まない文を非評判文と定義し、各文を評判文/非評判文に分類する2値分類タスクを考える。

品詞情報のみを用いた場合と意味類型情報を用いた場合とを比較するため以下に示す手法を用いて実験を行う。

#### 3.1 評判文抽出手法

まず、入力文をMeCab<sup>(3)</sup>を用いて形態素解析したあと以下に示す形態素解析の後処理を行う<sup>3</sup>。

<sup>3</sup>KNBコーパスは形態素アノテーション済みであるが、意味類型曖昧性解消器の入力がIPA品詞体系日本語辞書の品詞体系に基

- 直後に原形が「する」の動詞を伴うサ変接続の名詞は動詞として扱う
- 直後に助詞「を」と原形が「する」の動詞を続けて伴うサ変接続の名詞は動詞として扱う
  - － 上記2つに当てはまった場合、動詞「する」は用言として扱わない
- 直後に原形が「だ」「です」の助動詞を伴う形容動詞語幹の名詞は形容詞として扱う
- IPA品詞体系日本語辞書上の品詞細分類が自立でない動詞(「～している」の「いる」や「～し得る」の「得る」など)は用言として扱わない

この後処理はそれぞれ次に示す目的に基づく<sup>4</sup>。

- 意味類型曖昧性解消を「する」ではなく、語義を担うサ変名詞に対して行う
- 形容動詞(形容動詞語幹(名詞)+「だ/です」(助動詞))を形容詞として扱う
- 事実上機能語として振る舞う動詞を用言から除外する

その後、用言として扱われる動詞に対しては意味類型曖昧性解消器を用いて意味類型を付与し、形容詞に対しては意味類型の「形容」を常に付与する。

次に品詞情報、または意味類型情報を用いてコーパス中の全文から評判文の抽出を行いそれぞれの結果を比較する。先述のとおり評判表現の多くは形容表現が用いられていること、また、動作を表す表現は結果として客観的な事象を表していると考えられることから、以下に示す規則を用いて評判文の抽出を行う。

品詞情報のみを用いた規則 文中の最後に出現する用言の品詞が動詞であれば非評判文であると、形容詞であれば評判文とする。

意味類型情報を用いた規則 文中の最後に出現する用言の意味類型が動作であれば非評判文であると、変化、感覚・感情、形容のいずれかであれば評判文とする<sup>5</sup>。

づいていることを前提としているため再度形態素解析を行っている。

<sup>4</sup>なお、この後処理を行わずに品詞情報のみを用いて評判文抽出を行う場合、後述する結果と比較して悪い結果を示した。具体的には分類精度が0.574、F値が0.310となった。

<sup>5</sup>意味類型が変化、感覚・感情の場合に評判文とするのは予備実験の結果に基づく。

表 1: 品詞情報のみを用いた評判文抽出

正解\予測	評判文	非評判文	再現率 [%]
評判文	533	1512	26.1
非評判文	177	1964	91.7
適合率 [%]	75.1	56.5	59.7

評判文抽出の観点からの F 値=38.7%

表 2: 意味類型情報を用いた評判文抽出

正解\予測	評判文	非評判文	再現率 [%]
評判文	1442	603	70.5
非評判文	889	1252	58.5
適合率 [%]	61.9	67.5	64.4

評判文抽出の観点からの F 値=65.9%

文中の1つの用言だけを見るという単純な規則を用いるのは、結果に寄与する要因を減らし品詞情報と比較した意味類型情報の特徴を明らかにする意図に基づいている。また、最後の用言に注目しているのは、日本語文の主要部は後ろに出現しやすいという仮定に基づいている。

なお {品詞が動詞の用言} ⊃ {意味類型が動作の用言} であるので、意味類型を用いた場合は必ず評判文に対する再現率が上昇することに留意されたい。

### 3.2 結果・考察

表 1, 2 に品詞情報のみを用いた場合と意味類型情報を用いた場合の評判文抽出結果をそれぞれ示す。意味類型情報を用いた場合、品詞情報のみを用いた場合と比べて評判文/非評判文の分類精度において 4.7 ポイント、評判文抽出の観点から見た F 値において 27.2 ポイント上昇している。以下、各種誤りについて解析した結果について述べる。

品詞情報のみでは誤りで、意味類型情報を扱うことで正しく評判文であると判断された例は 909 文存在した。誤りには「～がわからなかったです。」「～と違います」「～と感じました」など、意味類型が感覚・感情の動詞で終わる文を多く含んでおり、動詞をその意味によって分類するという意味類型導入の意図が反映されており、有用性が確認できた。

品詞情報、意味類型情報どちらを用いた場合にも共通して誤りだった例は 780 文存在した。うち評判文を非評判文としてしまった例は 603 文、非評判文を評判文と判定してしまった例は 177 文だった。さらに前者について誤りの原因の解析を行った。全 603 文から 50 文を無作為に抽出した解析を行った結果の内訳は以下のとおりである。

意味類型が動作でも評判表現になる例 (16 件) 以下に示す表現においては下線部の動詞<sup>6</sup>の意味類型によらず評判表現になり得る。このような誤りはパターンマッチで個別に扱うことで解決できる。

<sup>6</sup>下線部が 1 動詞を表す。「～する」表記だがサ変動詞に限らない。

「～してしまう」「～れた」「～してくれる」「～しか～しない」「～してみたい」「～しやがって」「～しないでくれ」「～しない」「～すべきではないだろうか」「～して行きたい」

また、「普段から使っている」「観光客が押し寄せる」「騒ぎ散らす」のように特定の用法で評判表現になる例も存在した。

用言以外が評判表現 (のトリガ) になる例 (12 件) 用言以外の語にも評判表現になる、あるいはその語を伴うことではじめて評判表現になる、という語が存在する。具体的には「それ以上のもの」「宝もの」「人だかり」「バカ」「抜群」といった名詞、「ぜひ」「あっさり」「日常的に」といった副詞、「ありがとう」「ジーザス」といった感動詞、「超」といった接頭辞が確認された。「抜群」や「バカ」といった名詞に関しては形容動詞語幹として辞書に登録された方が良いと思われる。

評判表現が文の途中に出現する例 (11 件) 1 文に複数の節が含まれているなどの理由により、文の途中の用言が評判表現になっている場合がある。このような事例は係り受け解析や述語項構造解析を行い、その情報を用いることで扱うことができる。

形態素解析誤りに起因する誤り例 (7 件) うち 4 件はブログ記事の文体に起因するものであった (話し言葉、方言、小文字化、カタカナ化)。

形容動詞語幹の扱いに起因する誤り例 (3 件) 直後に「だ」や「です」を伴わない場合でも、形容表現として形容動詞語幹が用いられている場合がある。「～がお洒落。」のように文末に出現する場合が該当する。

意味類型付与誤りに起因する例 (1 件) 意味類型を付与する際に誤った意味類型が付与されてしまい、結果として誤った例が 1 件存在した。

表 3: KNB コーパス中の評判表現に含まれていた動詞「ある」の項

意味\項	名詞単体	名詞句
(属性を)有する	楽しみ, 問題, 雰囲気, 制限, 弱点, 開き, 魅力, おそれ, 価値, 歴史, 見所, ご利益, 種類, 威厳, 義務, 由来, メリット, 距離, 人気, チャンス, オチ, 使い道, 違和感, 文化, 割引, 安心感	例) 見直す必要, 得るもの, 参考になる点, QOL を高める要素, 興味深いもの, 引き込まれるもの, 歴史を感じさせるもの, 色々な機能
存在する	自然, 名所, 見所	例) いいところ, 有用なもの, 素敵なカフェ, お勧めスポット
(事象)が起こる	火事	例) 迷うこと, 無料になる場合, 不安になるとき
感情	興味, 文句, 思い入れ, 不満, 意見	例) 納得した覚え, そんなに使わなかった記憶, 面白いことになりそうという意識, 多いという印象

品詞情報のみでは正しかったが意味類型情報を扱うことによって本来は非評判文である文を評判文と判断し誤りとなった例は 712 文存在した。その中で最も大きい割合 (2 割程度) を占めていたのが, 文中の最後の用言が動詞「ある」の場合である。動詞「ある」は存在を表すので意味類型は常に形容であるが, 実際に評判表現になるか否かは項の名詞 (句) に依存する。例えば「レトロで綺麗なお庭があります。」は評判文であるが「~, その先に町家風の店内がある」は非評判文であった。表 3 に KNB コーパス中の評判表現に含まれていた動詞「ある」の項を示す。表中で「例」と書かれていない項目は, コーパス中で見つかった全ての項を挙げている。項が名詞単体の場合, 評判表現になるか否かは名詞自身の持つ極性によって決まると考える。「~がある」を伴った際の名詞の評価極性情報は日本語評価極性辞書 (名詞編)<sup>(4)</sup> に含まれており, これを利用することで評判表現か否かを判別することが可能になると考えられる。実際に表 3 中の名詞単体のうち「由来」「使い道」「オチ」「割引」「意見」を除いた全ての名詞が日本語評価極性辞書 (名詞編) に含まれていた。項が名詞句の場合は主辞である名詞に依存する場合と, 修飾部に依存する場合に分けられる。項の主辞が「機能」「印象」などの特定の名詞の場合は修飾部に依らず評判表現になりやすいが「もの」や「ところ」あるいは「カフェ」のような単体では評判に結びつきにくい名詞が主辞の場合は評判表現か否かが修飾部に依存すると考えられる。

## 4 まとめ

本研究では, 先行研究において日本語の用言を意味に基づいて分類することを目的として提案された意味類型の有用性を評判文抽出タスクで確認した。その結果, 品詞情報のみを用いた場合と比較して評判文/非評判文の分類精度で 4.7 ポイント, F 値で 27.2 ポイントの上昇を確認した。また誤り解析を通じて意味類型

情報だけでは扱うことのできない評判表現は, パターンマッチや項の名詞の属性情報を利用することで扱えるようになることを確認した。

意味類型情報は, これまで「動詞」「形容詞」という表層的な情報を用いてきた手法において品詞情報の代わりに用いることにより, 用言の意味をより正確に反映させることが可能になる。

## 使用した言語資源およびツール

- (1) IPA 品詞体系日本語辞書。Ver. 2.7.0。http://mecab.sourceforge.net/src/
- (2) 橋本力, 黒橋禎夫, 河原大輔, 新里圭司, 永田昌明。構文・照応・評判情報つきブログコーパスの構築。自然言語処理 Volume 18, Number 2, pp. 175-201。2011。
- (3) 工藤拓。形態素解析器 MeCab。Ver. 0.996。http://mecab.sourceforge.net/
- (4) 東山昌彦, 乾健太郎, 松本裕治。述語の選択選好性に着目した名詞評価極性の獲得。言語処理学会第 14 回年次大会論文集, pp. 584-587。2008。

## 参考文献

- [1] 中山匠, 山本和英。用言の新しい意味類型 - 作用性用言と形状性用言 -。言語処理学会 第 17 回年次大会 発表論文集, pp. 560-563, 2011。
- [2] Shohei Okada and Kazuhide Yamamoto。Semantic Type Disambiguation for Japanese Verbs。In *Proceedings of the IALP 2014*, pp. 6-9, 2014。
- [3] Thomas Scholz and Stefan Conrad。Opinion Mining in Newspaper Articles by Entropy-Based Word Connections。In *Proceedings of the EMNLP 2013*, pp. 1828-1839, 2013。
- [4] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一。意見抽出のための評価表現の収集。自然言語処理, Vol. 12, No. 3, pp. 203-222, 2005。
- [5] Tetsuji Nakagawa, Kentaro Inui, and Sadao Kurohashi。Dependency Tree-based Sentiment Classification using CRFs with Hidden Variables。In *Human Language Technologies: NAACL 2010*, pp. 786-794, 2010。
- [6] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts。Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank。In *Proceedings of the EMNLP 2013*, pp. 1631-1642, 2013。