

# 係り受け解析を用いた未知語意味カテゴリー推定手法の 有用性に関する諸検討

福岡 知隆<sup>†</sup> 服部 峻<sup>††</sup> 久保村千明<sup>†††</sup> 亀田 弘之<sup>††</sup>

<sup>†</sup> 東京工科大学大学院バイオ・情報メディア研究科 〒192-0982 東京都八王子市片倉町 1404-1

<sup>††</sup> 東京工科大学コンピュータサイエンス学部 〒192-0982 東京都八王子市片倉町 1404-1

<sup>†††</sup> 山野美容芸術短期大学美容総合学科 〒192-0396 東京都八王子市鎌水 530

E-mail: <sup>†</sup>g2110045e1@st.teu.ac.jp, <sup>††</sup>{hattori,kameda}@cs.teu.ac.jp, <sup>†††</sup>ckubomura@yamano.ac.jp

あらまし 近年、商品説明や観光地の案内などのサービスがコンピュータとの対話により行われることが増えてきている。しかし、それらサービスは未だ幾つかの問題を抱えており、その一つとして、対話中に発生する未知語の処理問題が存在する。この問題に対して我々はシステムが未知語の品詞や意味などの情報を自動的に推定することによるアプローチを試みる。本稿で提案する手法は未知語が含まれる入力文に対し係り受け解析器 CaboCha で係り受け解析を行って得られた文節とその係り受け関係、文中に存在する既知語の意味カテゴリーの類似性から取得した類似用例を用いて未知語の意味カテゴリー情報を推定する。その意味カテゴリー推定精度と提案手法が有効に作用する条件についての報告を行う。

キーワード 未知語, 未知語意味カテゴリー, 類似度

## Studies of Inference Method of Unknown Word Meaning Based on Syntactic Dependency and Its Usefulness

Tomotaka FUKUOKA<sup>†</sup>, Shun HATTORI<sup>††</sup>, Chiaki KUBOMURA<sup>†††</sup>, and Hiroyuki KAMEDA<sup>††</sup>

<sup>†</sup> Graduate School of Bionics, Computer and Media Sciences, Tokyo University of Technology

1404-1 Katakura-machi, Hachioji, Tokyo 192-0982, Japan

<sup>††</sup> School of Computer Science, Tokyo University of Technology

1404-1 Katakura-machi, Hachioji, Tokyo 192-0982, Japan

<sup>†††</sup> The General Department of Aesthetics, Yamano College of Aesthetics

530 Yarimizu, Hachioji, Tokyo 192-0396, Japan

E-mail: <sup>†</sup>g2110045e1@st.teu.ac.jp, <sup>††</sup>{hattori,kameda}@cs.teu.ac.jp, <sup>†††</sup>ckubomura@yamano.ac.jp

**Abstract** Recent developments of Information Technology enables us to talk with computers as new partners. In promoting products on the Web, guidance of a tourist resort, and so on. But there are many defects in the communication with computers. One reason is treatment of “unknown word”, which a computer doesn’t have as its knowledge in a database. To solve this problem we adopt a method that computer infer information about the unknown word automatically. For example POS sequence, semantic category, and so on. This paper proposes and evaluates a method to infer semantic category of an unknown word in an input sentence by syntactic dependency and semantic distance of words. Then report a condition that this inference method well affect. This method uses input phrase POS sequence and syntactic dependency by CaboCha and similarity of semantic category of the words to retrieve similar examples. And infer unknown word semantic category.

**Key words** unknow word, semantic word category, similarity degree

### 1. はじめに

近年の情報通信技術の進歩により、人間の対話相手が増えた。

人間に比べて膨大な情報の保持が可能なコンピュータである。チャットなどでの雑談相手、Web 上での商品の説明、介護における話し相手など、多岐にわたり人間はコンピュータと対話す

るようになった。しかし、人間同士の対話と比較すると、コンピュータの返答結果や対話の過程は劣っている場合が多い。その原因の一つが円滑性（発話者の意図に沿い、対話が速やかに行われること）の欠如である。コンピュータの単語辞書内に未登録な単語、即ち未知語に遭遇した場合にその現象は著しい。従来の対話システムでは未知語に対してまったく対応できなかったり、あるいはできたとしても人間への質問や話題転換が頻繁に起こってしまい、対話の円滑さが損なわれる場合がある。この問題を解決するため、対話システムにおける未知語処理を改善し人間とコンピュータ間の対話をより自然で円滑にする必要がある。一つの解決手法として、システムが自動的に未知語の情報を推定することで、既知語だけの発話と同様に応答することが可能になると考えられる。

本稿では上記の未知語情報の中で、未知語の意味カテゴリーを自動的に推定する手法を提案する。我々はすでに未知語とその前後の単語の品詞並びから入力文と類似する用例を検索し、未知語の意味カテゴリーを推定する手法の提案と評価を行った[1]。しかし、その推定結果は芳しいものではなかった。その理由の一つとして、未知語とその前後の単語の品詞並びをクエリとする類似用例検索条件が、正しい意味カテゴリーを推定するために必要な類似用例を検索するには厳しすぎたものと考えられる。

そこで本提案手法は、未知語の周辺情報に着目するのではなく、未知語を含む文の構造の類似性を重視する。未知語を含む入力文に対して係り受け解析を行い、得られた文節単位の品詞並びと文に含まれる意味カテゴリーを用いた、システムが保持する用例文との類似性に基づき推定する手法の提案、評価とその有効性の検討を行う。本稿での意味カテゴリーとは各単語の要素であり、シソーラスにおいて、類似した意味を持つ他の単語の意味カテゴリーと集合を形成する。例えば、晴天、雨天、曇天などの意味カテゴリーは天候を表す集合を形成するものとする。

## 2. 提案手法

本稿で提案する未知語意味カテゴリー推定手法では、あらかじめ用意した用例の中から入力文との類似用例を用いて未知語意味カテゴリーを推定する。類似用例は入力文と文節の品詞並び、単語の表層文字列、意味カテゴリーの類似度から決定する。

本提案手法は未知語を含む入力文との類似用例を検索するにあたり、文の構造が似ていることを重視する。ここで重視する文の構造とは文の主語と述語それぞれの周辺単語の品詞並びである。主語と述語それぞれの周辺の品詞をクエリとすることで、余計な修飾語にとらわれずに品詞並びが類似した用例を類似用例として検索することが可能であると考えられる。

システムはこの三つの類似性に基づいて類似した用例を一つ選択し、その用例の中で未知語に相当する単語から未知語の意味カテゴリーを推定する。

### 2.1 処理手順

本手法は以下の手順により、意味カテゴリー推定を行う。  
Step1: 意味カテゴリー情報付与 形態素解析器 MeCab を用

いて入力文を形態素解析する。得られた入力文の各形態素に対して、単語データベースを参照し意味カテゴリーを付与する。  
Step2: 入力文係り受け解析 類似用例を検索するクエリ生成に必要な情報を取得するため、係り受け解析器 CaboCha を用いて入力文の係り受け解析を行う。入力文の文節とその係り受け関係の情報を取得する。

Step3: クエリ生成 入力文の主語と述語をそれぞれ含む文節の品詞並びをクエリとして生成する。

Step4: 品詞並び類似用例検索 クエリにより入力文の類似用例を用例データベースから検索する。その際、検索された類似用例それぞれにおいて、未知語に相当する単語を入力文の未知語を含む文節の位置関係と文節の品詞並びから設定する。

Step5: 類似度計算 入力文と類似用例それぞれの単語の表層文字列と意味カテゴリーから二つの文の類似度を計算し、その類似度が最大の用例を選択する。

Step6: 意味カテゴリー推定 選択された類似用例の単語の中から入力文の未知語と品詞並びと同じ位置にある単語の意味カテゴリーを用いて未知語の意味カテゴリーを推定する。

### 2.2 データベース

本手法においては、単語データベースと用例データベースの二種類のデータベースを使用する。単語データベースは ipadic2.7.0 を使用し、各単語に関する意味カテゴリー情報は日本語 WordNet [2] から取得した。日本語 WordNet において意味カテゴリーを保持する品詞は名詞、動詞、形容詞、副詞の単語である。ただし、接頭詞、連体詞は形容詞に含まれている。また、意味カテゴリーは品詞毎に独立した関係性を持つ。

使用する品詞は ipadic の基本品詞 13 種類に加え、その中の助詞を細分化した 22 種類の品詞とした(表 1)。

表 1 助詞を細分化した品詞の分類

名詞	動詞	助動詞	形容詞
副詞	接続詞	感動詞	接頭詞
連体詞	記号	フィラー	その他
格助詞	副助詞	終助詞	係助詞
接続助詞	並立助詞	副詞化	連体化
副助詞 / 並立助詞 / 終助詞			特殊

用例データベースは、Web 上に公開されている対話コーパス [3] の中で二人の対話用例を用いて作成する。対話文章を文単位に分割し、一つのレコードとしてデータベースに格納する。コーパスから用例を獲得する際には単語データベースに存在しない未知語が出現した文と、形態素解析を行い得られた文中の単語の中で名詞、動詞、形容詞、副詞、接頭詞、連体詞の単語が意味カテゴリーを持たない文を除く。結果、用例データベースは 9233 の用例を保持している。

### 2.3 意味カテゴリー情報付与

入力文を MeCab により形態素単位に分割し、それぞれの単語をクエリとし、単語データベースに未知語が否か、また意味カテゴリー情報を問い合わせる。前述の単語データベースの仕様から、意味カテゴリーを持つことができるのは名詞、動詞、形容詞、副詞、接頭詞、連体詞のいずれかの単語だけである。

また、意味カテゴリー情報は上記の品詞の単語のすべてが保持しているわけではない。

## 2.4 入力文係り受け解析

未知語を含む入力文と類似した用例を検索するため、入力文の主語と述語をそれぞれ含む文節を取得する。文の主体である主語とその述語部分の品詞並びを用いることで、文の修飾詞を除去した基本的な構造が類似する用例を検索する。

CaboCha により入力文を文節単位に分割し、また各文節の係り受け関係の情報を取得する。各文節は先頭から番号が振られ、また自身の係り先の文節番号を保持しており、係り先がない文節は - 1 という値を持つ。これら各文節の係り先番号はどの文節から辿り始めても、係り先がない文節に辿りつく。

そこで文の基本的な部分を表していると考えられる係り先の文節番号が - 1 である文節とその文節番号を係り先文節番号として保持する文節の品詞並びを類似用例の検索クエリに用いる。なお、得られたすべての文節に対し、複合名詞に対応するため、名詞が二つ以上連続した場合は一つの名詞に変換し、また、語の活用形などを除去するため、各文節の先頭から二形態素以外は削除する。

## 2.5 クエリ生成

係り受け解析結果を用いて類似用例検索クエリとなる品詞並びを生成する。係り受け解析の結果、三つ以上の文節が条件に当てはまった場合は、その内の二つの文節を選択し、すべての組み合わせにおける品詞並びをクエリとする。未知語がクエリとなる文節内に存在する場合は、未知語の品詞をワイルドカードとし、必ずその文節をクエリに用いる。未知語がクエリとなる文節内に存在しない場合は、クエリとなる文節の間に存在するか否かを記憶する。例として図 1 では最後の文節が係り先の無い文節となり、「天気は」「よい」「晴れます。」の三つの文節が使用され、二通りのクエリが生成される。もし、「天気」を未知語とした場合、未知語を含む文節を必ずクエリに用いるため、「天気は」「晴れます。」の二文節がクエリ生成として使用される。

入力文: 東京の天気は気持ちのよい晴れです。

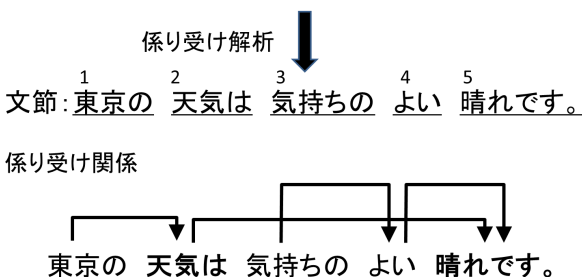


図 1 係り受け解析結果一例

## 2.6 類似用例検索

生成したクエリを用いて、入力文の品詞並びと類似する用例を用例データベースから検索する。また、それぞれの類似用例において、未知語に相当する単語を決定する。

未知語がクエリとなる文節に存在する場合、検索された類似用例において、クエリとなるもう一つの文節との並び順から判断される文節に対して、2.4 で加工した未知語を含む文節の未知語を除く品詞並びが一致する文節を選択し、その文節の先頭の単語を類似用例における未知語に相当する単語とする。未知語がクエリ内に存在しない場合は、入力文における未知語が存在する文節とクエリとなる文節の並び順から判断される文節に対して、同様に未知語に相当する単語を選択する。この際、選択された単語が意味カテゴリーを持たなかった場合、その用例を類似用例から除外する。

## 2.7 類似度計算

類似用例に対して類似度を計算し、最も値の大きい最類似用例の一つを選択する。類似度は単語の表層文字列の類似と意味カテゴリーの類似を用いて計算する。

単語の表層文字列を用いた類似度の計算式は以下のダイス係数を用い、 $w_s(x, y)$  を類似度、 $S(x)$  を入力文  $x$  の単語の集合、 $S(y)$  を類似用例  $y$  の単語の集合とする。一つの文に二つ以上の同じ単語が存在した場合は、それぞれを異なる要素として扱う。

$$w_s(x, y) = 2 \times \frac{|S(x) \cap S(y)|}{|S(x)| + |S(y)|} \quad (1)$$

次に意味カテゴリーの類似度を計算する。意味カテゴリー間の類似度を計算し、その計算結果を用いて文の意味カテゴリーの類似度を計算する。意味カテゴリー間の類似度は、日本語 WordNet で定義されている意味概念木とその上で定義される単語間の距離に基づき計算される。意味カテゴリー一つにつき、日本語 WordNet の「上位」関係に基づいた意味カテゴリーの木構造を用いて各単語からその最上位意味カテゴリーまでの上位意味カテゴリー群の塊を一つの要素とする。単語によっては複数の意味カテゴリーを持つ場合や同じ最上位意味カテゴリーまでに含まれる上位意味カテゴリーが異なる場合があるが、それらはすべて異なる要素として扱う。システムはこれらの要素を用いて意味カテゴリー距離を取得し、意味カテゴリー間の類似度を求める。ここでの意味カテゴリー距離とは、日本語 WordNet の意味カテゴリーの「上位」関係に基づく木構造においての二つの意味カテゴリー間の枝の数である (図 2)。

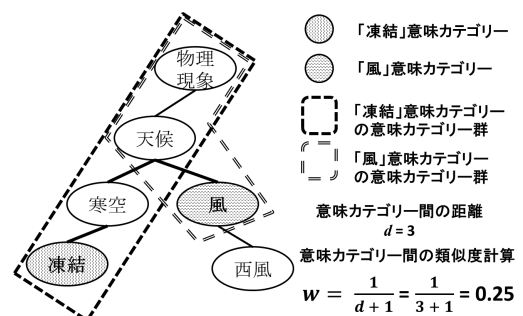


図 2 意味概念木上の意味カテゴリーの位置と距離

意味カテゴリー  $a, b$  の類似度は以下の式で計算される。 $w(a, b)$  を類似度とし、 $d(a, b)$  は意味カテゴリー間の最小の枝の数である ( $0 < w(a, b) \leq 1$ )。なお、木構造上に存在しない意

意味カテゴリーとの類似度は0となる。

$$w(a, b) = \frac{1}{d(a, b) + 1} \quad (2)$$

入力文と用例の類似度を計算する。この計算手法は類似度が大きい意味カテゴリーの組み合わせから順に選択する。以下の順に行われる。

Step1 入力文  $x$  に含まれる意味カテゴリーの集合  $C(x)$  と用例  $y$  に含まれる意味カテゴリーの集合  $C(y)$  それぞれの組み合わせの意味カテゴリーの類似度を計算する。

Step2 最も類似度  $w(a, b)$  が大きい意味カテゴリー  $a$  と  $b$  の組を一つ選択する。類似度が最大の組が複数あった場合は  $C(y)$  において先に出現する意味カテゴリーを含む組を優先する。 $C(y)$  の意味カテゴリーが同じであった場合は、 $C(x)$  において先に出現する意味カテゴリーを含む組を選択する。

Step3 選択された組の意味カテゴリーを含むその他の組の類似度を  $-1$  に変更する。

Step4 Step2, Step3 を入力文  $x$  と類似用例  $y$  それぞれの意味カテゴリー数の最小値と同じ回数繰り返す。

二つの文の類似度  $w_c(x, y)$  を計算する(式(3))。

$$w_c(x, y) = \frac{\sum_{a \in C(x)} \sum_{b \in C(y)} w(a, b)}{\sqrt{|C(x) \times C(y)|}} \quad (3)$$

## 2.8 意味カテゴリー推定

類似用例の中から意味カテゴリー類似度が最も大きい用例を一つ選択する。意味カテゴリー類似度が同一の用例が複数存在した場合は、表層文字列の類似度が大きい用例を選択し、表層文字列の類似度も同じ場合は、先に用例データベースに格納している用例を選択する。選択した最類似用例の未知語に相当する単語の意味カテゴリーを未知語の意味カテゴリーとする。

## 3. 評価

提案した未知語意味カテゴリー推定手法の推定精度の評価を行った。比較対象として、未知語の前後の品詞並びをクエリとする類似用例検索結果を用いた意味カテゴリー推定手法についても評価を行った。

### 3.1 評価方法

提案した未知語意味カテゴリー推定手法の推定精度を評価するため、未知語を含む一つ入力文を複数用意し、提案手法による意味カテゴリー推定を行い、その推定精度を評価する。

#### 3.1.1 推定精度計算方法

意味カテゴリー推定精度の計算は推定された未知語の意味カテゴリーの集合があらかじめ用意した未知語の正答意味カテゴリーの集合に対して、どの程度の割合で同じ意味カテゴリーが含まれているかを計算することで行う。意味カテゴリー推定精度を推定された未知語  $w$  の意味カテゴリーの集合を  $C(w)$ 、予め用意した未知語の正答意味カテゴリーの集合を  $C'(w)$  とした以下の式により計算する。この値が大きいほど、その意味カテゴリー推定結果は正しい確率が大きいものとする。

$$2 \times \frac{|C(w) \cap C'(w)|}{|C(w)| + |C'(w)|} \quad (4)$$

#### 3.1.2 未知語を含む入力文生成

形態素解析器 MeCab においては、既知語であってもモカナ表記することで未知語と判断される場合が多い。本稿ではこれを利用し、既知語をカナ変換することで、正答となる意味カテゴリーが明確な未知語とその用例を作成する。

WordNet に保存されている用例に対して形態素解析、係り受け解析を行う。単語を取り出しそれぞれの意味カテゴリーを単語データベースから取得し、用例の文節区切りとその品詞並びを取得する。ただし、三つ以上の単語が意味カテゴリーを持たない用例は除く。意味カテゴリー情報を持つ単語をカナ変換し、再度その用例を形態素解析、係り受け解析を行い、文節区切り、品詞並びが変化せず、文の構造が変化していない用例を未知語を含む入力文として取得する。カナ変換は、一単語ずつ行う。また、同一の単語が二つ以上存在する場合は、その単語をカナ変換した用例は用いない。この手法により 32056 文(未知語の重複あり)の入力文を作成した。生成した入力文セットにおける未知語の正答品詞別の入力文数を表 2 に示す。

表 2 品詞別の未知語入力文数

名詞	動詞	形容詞	副詞
26441	3677	1034	904

### 3.2 推定結果

提案手法における類似用例検索を行った時点での結果を図 3 に、未知語の前後の品詞並びによる類似用例検索結果を図 4 に示す。また、両手法における品詞別の平均類似用例検索数を図 5 に示す。

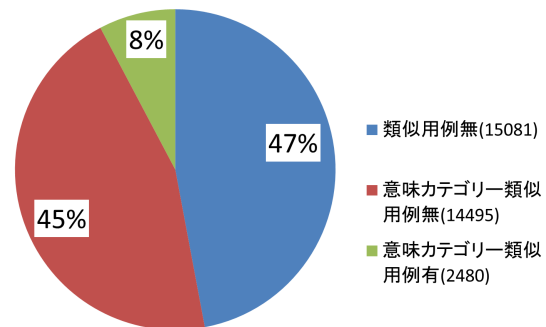


図 3 提案手法類似用例検索結果

提案手法の主語と述語をそれぞれ含む文節の品詞並びにより類似用例の検索を行った結果、入力文の 47% は品詞並びの類似用例が存在せず(類似用例無)、45% は品詞並びの類似用例は存在するが、未知語の該当単語における意味カテゴリーが類似する用例が存在せず(意味カテゴリー類似用例無)、8% が未知語の正答意味カテゴリーの集合と一つ以上一致する未知語該当単語を含む類似用例が存在する(意味カテゴリー類似用例)

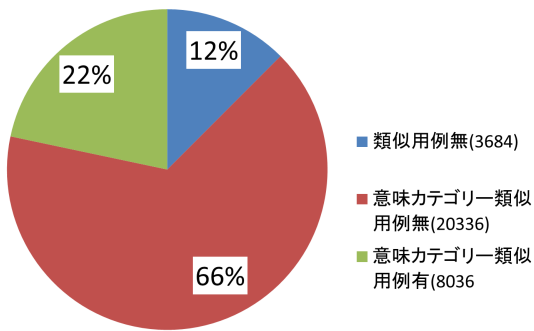


図 4 未知語の前後の品詞並びによる類似用例検索結果

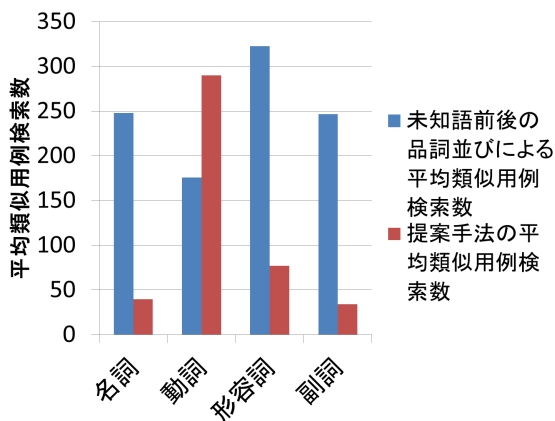


図 5 二つの手法の平均類似用例検索数

有)。つまり 9 割以上の確率で、正しい未知語意味カテゴリ推定が必ず行われない結果となる。未知語の前後の品詞並びによる類似用例の検索結果、入力文の 12 % は類似用例が存在せず、66 % は品詞並びの類似用例は存在するが、未知語の該当単語における意味カテゴリが類似する用例が存在せず、22 % が未知語の正答意味カテゴリの集合と一つ以上一致する未知語該当単語を含む類似用例が存在する。

二つの手法における品詞別の類似用例検索結果を表 3, 4 にそれぞれ示す。

表 3 提案手法の類似用例検索結果

	名詞	動詞	形容詞	副詞
類似用例無	13165	971	492	453
意味カテゴリ類似用例無	12194	1439	472	390
意味カテゴリ類似用例有	1082	1267	70	61

表 4 未知語前後の品詞並びによる類似用例検索結果

	名詞	動詞	形容詞	副詞
類似用例無	3303	223	93	65
意味カテゴリ類似用例無	17405	1855	573	503
意味カテゴリ類似用例有	5733	1599	368	336

提案手法による類似用例検索を行った結果、正答品詞が動詞

である場合が、最も高い比率で未知語の正答意味カテゴリの集合と一つ以上一致する未知語該当単語を含む類似用例を検索した。未知語前後の単語の品詞並びに類似用例検索では、正答品詞が名詞である場合が、最も高い比率で未知語の正答意味カテゴリの集合と一つ以上一致する未知語該当単語を含む類似用例を検索した。

二つの手法においての意味カテゴリ類似用例無、意味カテゴリ類似用例有の場合の類似用例と意味カテゴリ推定成功の最類似用例の意味カテゴリ類似度の平均値を表 5, 6 に示す。

表 5 提案手法の類似用例検索結果の平均意味カテゴリ類似度

	名詞	動詞	形容詞	副詞
意味カテゴリ類似用例無	0.22	0.26	0.20	0.21
意味カテゴリ類似用例有	0.32	0.34	0.37	0.35
意味カテゴリ推定成功	0.27	0.39	0.36	0.52

表 6 未知語前後の品詞並びによる類似用例検索結果の平均意味カテゴリ類似度

	名詞	動詞	形容詞	副詞
意味カテゴリ類似用例無	0.30	0.20	0.25	0.25
意味カテゴリ類似用例有	0.38	0.31	0.38	0.33
意味カテゴリ推定成功	0.33	0.36	0.58	0.36

二つの手法において未知語意味カテゴリ推定を行った結果、少なくとも一つ以上の未知語の正答意味カテゴリを推定できた意味カテゴリ推定成功入力文においての品詞別に入力文数とその平均意味カテゴリ推定精度を表 7, 8 に示す。

表 7 提案手法の正答品詞別意味カテゴリ推定精度

	名詞	動詞	形容詞	副詞
推定成功入力文数	43	36	1	2
平均意味カテゴリ推定精度	0.66	0.55	1.0	0.61

表 8 未知語前後の品詞並びを用いる手法の正答品詞別意味カテゴリ推定精度

	名詞	動詞	形容詞	副詞
推定成功入力文数	204	39	3	16
平均意味カテゴリ推定精度	0.67	0.70	0.59	0.54

提案手法により、少なくとも一つ以上の正答意味カテゴリを推定できた意味カテゴリ推定成功入力文は 82 文である。入力文に対する意味カテゴリ推定成功時の推定精度は平均で 71 % の推定精度となるが、すべての入力文に対する推定精度では約 0.2 % であった。未知語前後の単語の品詞並びを用いた手法では 262 文が意味カテゴリ推定成功となり、推定成功時の推定精度は平均で約 63 % であり、すべての入力文に対する推定精度は約 0.5 % であった。

## 4. 考 察

以下に類似用例検索結果，意味カテゴリー類似度，意味カテゴリー推定精度についての考察を記す．なお，正答品詞が形容詞，副詞となる未知語の意味カテゴリー推定成功入力文数は少ないので，その推定精度に対する考察は行わない．

### 4.1 類似用例検索

本提案手法における類似用例は形態素数を少数に限定した主語と述語をそれぞれ含む二つの文節の品詞並びの組み合わせを用いることで，文の構造が類似した用例を検索することを意図した．しかし，図 3~5，表 3，4 の結果から今回提案した手法では類似用例の検索結果の改善が行われず，逆に類似用例検索数が低下している．

過去の研究 [4] において，未知語を含む連続した形態素の品詞並びをクエリとする場合は未知語とその前後の形態素を用いる類似用例検索では，未知語とその前後一，二文字の形態素による結果が良好なものであったことに鑑みると，本提案手法の四形態素の品詞並びをクエリとする点が失敗の原因とは考えにくい．よって，その原因は形態素数ではなく，文の主語と述語をクエリに含めた点であると考えられる．

主語と述語をそれぞれ含む文節の品詞並びをクエリとしているため，場合によっては検索した用例文中には未知語が存在する文節と同じ位置に文節が存在しない．このような検索結果は除くため，本提案手法における類似用例検索ではクエリである品詞並びに加え，用例中に未知語が存在すると予想される文節を持つか否かも条件として加味しなければならないため，未知語前後の品詞並びによる類似用例検索よりも検索条件が厳しくなっており，類似用例検索数が少なくなっていると考えられる．

### 4.2 意味カテゴリー類似度

表 3，4 の結果から，検索された類似用例と最類似用例それぞれの類似度について提案手法による改善は行われていない．また，意味カテゴリー推定が成功した最類似用例の類似度の平均値が，意味カテゴリー類似用例有の類似用例の平均値に比べてあまり変化がないことから，類似度の高さの意味カテゴリー推定の成功率との関係性が低いといえる．これは今回設定した意味カテゴリー類似度は意味カテゴリー推定が成功する類似用例の絞り込みを効果的に行えていないともいうことができる．その原因として本稿で定義した意味カテゴリー類似度は品詞が異なる単語同士の意味カテゴリーの類似度を 0 するため，品詞別の意味カテゴリー数が異なるとその正確性が損なわれるためだと考えられる．

### 4.3 意味カテゴリー推定精度

表 7，8 の結果から，本提案手法による意味カテゴリー推定が成功する場合は 60 % 程度の精度で意味カテゴリー推定が行えるが，意味カテゴリー推定が行える件数が少ないため，全ての入力文においての平均意味カテゴリー推定精度は 1 % を下回る結果となっている．これは類似用例検において，類似用例の取りこぼしが多く，意味カテゴリー類似度による最類似用例絞り込みが正しく有効でないためである．

## 5. おわりに

本稿では上記未知語情報の中で，未知語の意味カテゴリーを自動的に推定する手法を提案し，その意味カテゴリー推定精度の評価と有効な利用条件についての検討を行った．その結果，主語と述語をそれぞれ含む文節の品詞並びによる類似用例の検索とその結果を用いた意味カテゴリー推定は有効ではないという知見と意味カテゴリーを類似度を計算する上で，品詞毎のソーラスの木構造を考慮する必要があるという知見を得た．

### 文 献

- [1] 福岡 知隆, 服部 峻, 久保村 千明, 亀田 弘之. “単語間の意味カテゴリー距離に基づく用例ベース未知語意味カテゴリー推定.” 第 10 回 情報科学技術フォーラム (FIT'11), 4F-4. (2011).
- [2] 日本語 WordNet, <http://nlpwww.nict.go.jp/wn-ja/> (2011).
- [3] 名大会話コーパス, <http://dbms.ninjal.ac.jp/nuc/index.php> (2011).
- [4] 福岡 知隆, 服部 峻, 久保村 千明, 亀田 弘之: “品詞並び検索条件の段階的緩和による用例ベース未知語品詞推定,” 第 90 回 人工知能学会 知識ベースシステム研究会 (SIG-KBS), 人工知能学会研究会資料, SIG-KBS-B001-04, pp.23-30, (2010).